IN THE UNITED STATES DISTRICT COURT
FOR THE WESTERN DISTRICT OF PENNSYLVANIA

UNITED STATES OF AMERICA	)	
	)	
v.	)	Criminal No. 18-292
	)	
ROBERT BOWERS	)	

MOTION TO EXCLUDE DNA EVIDENCE, AND REQUEST FOR *DAUBERT* HEARING AND FOR DISCOVERY IN SUPPORT OF MOTION

1001 Liberty Avenue
Suite 1500
Pittsburgh, PA 15222
(412) 644-6565

Judy Clarke
Clarke Johnston Thorp & Rice, PC

Michael J. Novara
First Assistant Federal Public Defender

Elisa A. Long
Supervisory Assistant Federal Public Defender

Attorneys for Defendant,
Robert Bowers

Table of Contents

IN THE UNITED STATES DISTRICT COURT
FOR THE WESTERN DISTRICT OF PENNSYLVANIA

UNITED STATES OF AMERICA	)	
	)	
v.	)	Criminal No. 18-292
	)	
ROBERT BOWERS	)	

MOTION TO EXCLUDE DNA EVIDENCE, AND REQUEST FOR *DAUBERT* HEARING AND FOR DISCOVERY IN SUPPORT OF MOTION

Defendant Robert Bowers, by counsel, moves the Court for an order, following an evidentiary hearing, excluding the government's DNA evidence at the trial of this case. Mr. Bowers makes this motion pursuant to Federal Rules of Evidence Rules 104(a), 402, 403, 702, and 703; Federal Rule of Criminal Procedure Rule 16; 18 U.S.C. § 3593(c); and the Fifth, Sixth, and Eighth Amendments to the United States Constitution.

The grounds for this motion are that:

- there is no reliable scientific basis for the specific type of DNA testimony being offered in this case and thus this testimony is inadmissible under *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), and *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999);
- the testimony is inadmissible under the 2000 amendments to Rule 702 and Rule 703 in that (a) the testimony is not based upon sufficient facts or data, (b) the testimony is not the product of reliable principles and methods, and (c) the individuals who performed the DNA examinations in this case have not applied the principles and methods reliably to the facts of the case;

- the specific testimony based on DNA testing that is being offered in this case is irrelevant and inadmissible under Rule 402; and
- any weak probative value of the proposed testimony is also substantially outweighed by the danger of unfair prejudice, confusion of the issues, and misleading the jury, and by considerations of undue delay, waste of time, and needless presentation of cumulative evidence and is thus inadmissible under Rule 403 and the due process, fair trial, and cruel and unusual provisions of the Constitution and 18 U.S.C. § 3593(c).

In addition, and preliminary to ruling on the admissibility of the DNA evidence, Mr. Bowers asks this Court to compel the Government to provide forthwith the "case litigation" packets that support the findings and conclusions of FBI DNA Casework Analyst Marcy L. Plaza's 16-page DNA Report, dated January 23, 2019, which was provided to the defense on March 19, 2019 ("DNA Report").[1] The case litigation packets, as well as additional information that the defense may need to request after an initial review of the litigation packets, are essential for the defense to determine, through independent expert analysis, whether there are grounds in addition to those raised in this motion for exclusion of the government's DNA testing results and of the results of the probabilistic genotyping software utilized to generate the DNA statistical analysis in this case. Identifying the proper experts and narrowing the *Daubert* issues require the production of the case litigation packet materials.

[1] The DNA Report is attached as Exhibit 1.

At the status hearing held on November 2, 2020, the defense noted the complexity of *Daubert* issues surrounding the DNA evidence and explained that this was an area where the defense could make progress in this case despite the COVID delays. Transcript of November 2, 2020 Telephone Conference at 4-5. The defense explained that the government's provision of the case litigation packets would facilitate the *Daubert* litigation, and the Court indicated that it would entertain a written request to that effect. *Id.* at 7–8, 19 ("I'm going to wait and see what the defense files in terms of the *Daubert* motions.").

In accordance with the Court's instructions, Mr. Bowers brings the present motion, which outlines a number of *Daubert* issues that can be raised in this case based on Analyst Plaza's conclusory report alone. Necessarily, however, this preliminary *Daubert* motion is not exhaustive and must be supplemented in the future since any flaws in the underlying testing results and performance of the DNA and STRmix™ technology are impossible to fully understand without access to the actual testing data, laboratory materials, and other pertinent information that the defense may need to request after review of those materials.

I. INTRODUCTION

"The past two-and-a-half decades have seen tremendous growth in the use of DNA evidence in crime scene investigations as well as paternity and genetic genealogy testing." John Butler, FUNDAMENTALS OF FORENSIC DNA TYPING 5 (2010) ("Butler,

Fundamentals").[2] However, as with all of the forensic sciences, "[i]t is important to keep in mind that forensic DNA tests must be performed carefully in order to obtain reliable results." *Id.* at 7. Stringent standards and precise protocols must be "faithfully followed by forensic DNA laboratories to help maintain a high level of quality in the data obtained" and "information produced as part of a forensic examination must be reliable and able to hold up in court under rigorous scrutiny." *Id.*

In their 2009 landmark report, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD (2009) ("NRC 2009 Report"), the National Academy of Sciences, National Research Council's Committee on Identifying the Needs of the Forensic Science Community, reminds us:

> Among existing forensic methods, only nuclear DNA analysis has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between an evidentiary sample and a specific individual or source. Indeed, DNA testing has been used to exonerate persons who were convicted as a result of the misapplication of other forensic science evidence. However, this does not mean that DNA evidence is always unassailable in the courtroom. There may be problems in a particular case with how the DNA was collected, examined in the laboratory, or interpreted, such as when there are mixed samples, limited amounts of DNA, or biases due to the statistical interpretation of data from partial profiles.

NRC 2009 Report at 100.[3]

---

[2] John Butler's textbooks on forensic DNA analysis—including Fundamentals of Forensic DNA Typing (2010), Forensic DNA Typing: Methodology (2012) ("Butler, DNA Methodology"), and Advanced Topics in Forensic DNA Typing: Interpretation (2015) ("Butler, Advanced Topics")—are considered authoritative in the field of forensic DNA analysis. They are widely used in forensic science education, as well as cited in laboratories' protocols. They have also been relied upon by the Third Circuit. *See, e.g.*, *United States v. Grinnage*, 486 F. App'x 325, 330 (3d Cir. 2012). Either electronic or hard copies of these books will be provided to the Court upon request.

[3] Available at https://www.nap.edu/catalog/12589/strengthening-forensic-science-in-the-united-states-a-path-forward (last visited Jan. 3, 2021). As the Court is probably aware, the National Academy of Sciences (NAS) is a private, non-profit society of distinguished scholars. Established by an Act of Congress, which was signed by President Abraham Lincoln in 1863, the

The reason why DNA evidence must be subjected to rigorous scrutiny is obvious but bears emphasis. "Testimony from forensic DNA analysts can be overwhelming to jury members—who often have limited exposure to the concepts of biology, genetics, and the technology used to generate DNA profiles." Butler, DNA Methodology at 542. "There is a considerable aura to DNA evidence. Because of this aura it is vital that weak evidence is correctly represented as weak or not presented at all." *Id.* (internal quotation omitted). As emphasized by the Supreme Court, "[g]iven the persuasiveness of [DNA] evidence in the eyes of the jury, it is important that it be presented in a fair and reliable manner." *McDaniel v. Brown*, 558 U.S. 120, 136 (2010).

Unfortunately, jurors' and courts' heavy reliance on DNA experts can be woefully

---

NAS is charged with "providing independent, objective advice to the nation on matters related to science and technology." NAS Mission Statement (available at http://www.nasonline.org/about-nas/mission/) (last visited Jan. 3, 2021). In 1916, the Academy established the National Research Council at the request of President Wilson to recruit specialists from the larger scientific and technological communities to participate in the Academy's advisory work to the government. The NAS 2009 Report, therefore, constitutes an assessment not by individual scientists or scholars but by a scientific institution, the first such assessment of forensic science evidence by a mainstream scientific institution of any kind.

The NAS has published two other books on the admissibility of DNA evidence. *See* National Research Council, DNA TECHNOLOGY IN FORENSIC SCIENCE (1992) (hereinafter"NRC I"); National Research Council, THE EVALUATION OF FORENSIC DNA EVIDENCE (1996) (hereinafter "NRC II"). Both are available online at https://www.nap.edu/search/?term=DNA+Technology+in+Forensic+Science+&x=20&y=16 and https://www.nap.edu/search/?rpp=20&ft=1&term=The+Evaluation+of+Forensic+DNA+Evidence+ (last visited Jan. 3, 2021).

One or more of these three books are relied upon extensively in most judicial opinions dealing with the admissibility of forensic DNA evidence, although the first two are now significantly out of date. In other words, when the NAS speaks, courts listen. *See, e.g*, *Melendez-Diaz v. Massachusetts*, 557 U.S. 305, 318 (2009) (quoting the NRC 2009 Report for the propositions that "[f]orensic evidence is not uniquely immune from the risk of manipulation" and that "'[t]he forensic science system, encompassing both research and practice, has serious problems that can only be addressed by a national commitment to overhaul the current structure that supports the forensic science community in this country.'").

misplaced. It is assumed, for instance, even by trained judges, and certainly therefore by most jurors, that each new DNA typing kit that comes on the market is merely another in a series of improved ways to apply long-accepted science. *See, e.g.*, *United States v. Trala,* 162 F. Supp. 2d 336, 348 (D. Del. 2001). But what gets overlooked with such a perspective is the very real downside of scientific development in this rapidly changing field of forensic science. As explained by one of the leading researchers and practitioners in this field:

> In recent years, the sensitivity of DNA profiling has steadily increased, so that now the analysis of just a "handful" of cells is not only possible, but also routine in most forensic laboratories. This has been possible by the introduction of new multiplexes by manufacturers (these are biochemical systems used to detect DNA profiles), along with new highly sensitive detection platforms.
>
> However, as the sensitivity of DNA profiling technology increases, there is a parallel increase in the uncertainty of associations. This is because DNA is everywhere in the environment. It can be transferred passively, e.g., by touching a surface, or by secondary transfer, mediated by a person other than the defendant. DNA will persist indefinitely in a dry enviornment, hence there is no implicit information attached to the DNA profile that gives a clue to the "how" and "when" transfer occurred.

Peter Gill, MISLEADING DNA EVIDENCE: REASONS FOR MISCARRIAGES OF JUSTICE xii (2016) ("Gill, Misleading DNA").

As Dr. Gill and others have demonstrated, forensic DNA testing is subject to misinterpretation, faulty analysis, mislabeling, contamination, and other serious deficiencies. *See also* NRC 2009 Report at 47 ("[A]lthough DNA analysis is considered the most reliable forensic tool available today, laboratories nonetheless can make errors working with either nuclear DNA or mtDNA—errors such as mislabeling samples, losing

samples, or misinterpreting the data."). Also, the tests themselves may be unreliable, skewed or misinterpreted by human error. As was stated in *United States v. Bentham*, 414 F. Supp. 2d 472, 473 (S.D.N.Y. 2006), "[f]alse positives—that is, inaccurate incriminating test results—are endemic to much of what passes for 'forensic science.' . . . Even the 'gold standard' of forensic testing, DNA tests, may, because of human error, prove fallible." *See also United States v. Grinnage*, 486 F. App'x 325, 330 (3d Cir. 2012) ("LCN [low copy number] testing might be less reliable than the typical PCR/STR method, especially where the proponent of the evidence seeks to prove a negative. This is because stochastic events can lead to phenomena such as 'allele dropout,' causing part of a profile to disappear. *See generally* John M. Butler, FORENSIC DNA TYPING 68, 167–70 (2d ed. 2005). Consequently, there may be cases where a *Daubert* hearing will be necessary to discern a minimum acceptable mass threshold, below which the PCR/STR methodology is unreliable."); *In re Paoli R.R. Yard PCB Litig.*, 916 F.2d 829, 853 (3d Cir. 1990) ("[T]he district court must have a proper and reviewable foundation for making its admissibility findings.").

In this particular capital case, in light of the issues discussed below, this Court should grant the requested discovery and then, after the defense has had the opportunity to review the discovery and obtain any necessary additional discovery, hold an evidentiary hearing in order to properly and fully exercise its gate-keeping function under *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 590 (1993), Federal Rules of Evidence Rules 403, 702, and 703, and 18 U.S.C. § 3593(c) to ensure that unreliable,

prejudicial, confusing, and misleading DNA evidence does not reach the jury.[4] The need to do so is particularily acute in this death penalty case where, under the Eighth Amendment, there is a need for enhanced reliability at all phases of the case, and where countless cases have now demonstrated that despite its widespread use, DNA evidence is by no means failsafe.

In fact, when DNA testing was first introduced, experts argued that a false match (a test result that incriminated an innocent defendant) was impossible. *See* William C. Thompson, Franco Taroni, & Colin G.G. Aitkin, *How the Probability of a False Positive Affects the Value of DNA Evidence,* 48 J. FORENSIC SCI. 1, 2 (2003) [hereinafter "Thompson, False Positive"]. "This claim is now broadly recognized as wrong in principle , and it has repeatedly proven wrong in practice." *Id.* at 2. False matches and flawed DNA test results can cause wrongful convictions of innocent people, and DNA

[4] Section 3593(c) states in relevant part that at the sentencing phase of a federal capital case:

> Information is admissible regardless of its admissibility under the rules governing admission of evidence at criminal trials except that information may be excluded if its probative value is outweighed by the danger of creating unfair prejudice, confusing the issues, or misleading the jury.

Under this statute, at a capital sentencing trial, district courts "retain the discretion to exclude any type of unreliable or prejudicial evidence." *United States v. Pepin*, 514 F.3d 193, 204 (2d Cir.2008). Further, "Section 3593(c) provides the district court with greater discretion to exclude unfairly prejudicial or confusing information than the district court has during the guilt phase. This is so because § 3593(c) requires only that the countervailing interests 'outweigh' the information's probative value, while Federal Rule of Evidence 403 permits exclusion only where the countervailing interests 'substantially outweigh[ ]' the evidence's probative value." *United States v. Lujan*, 603 F.3d 850, 854 (10th Cir. 2010).

evidence has been called into question or proven faulty in several cases around the country.[5]

Errors resulting from misinterpreted, mislabeled, or contaminated DNA samples "appear to be chronic and occur even at the best DNA labs." William C. Thompson, *Tarnish on the 'Gold Standard;' Understanding Recent Problems in Forensic Testing*, THE CHAMPION, Feb. 2006, at 10–12 [hereafter "Thompson, Tarnish on the Gold Standard"]; D. Syndercombe Court, *DNA analysis: Current Practice and Problems*, in John Gall and Jason Payne-James, ed., CURRENT PRACTICE IN FORENSIC MEDICINE 208 (2011) ("DNA evidence is often misinterpreted by judge and jury alike and there are a number of well-documented errors in presenting DNA evidence in court that have led to successful appeals."). Significant errors have been documented in California, Maryland, Minnesota, Nevada, North Carolina, Pennsylvania, Texas, and Washington.[6] In fact, there

[5] *See, e.g.*, *People v. John*, 52 N.E.3d 1114, 1126 (N.Y. 2016) ("We note that the amicus curiae brief by the Innocence Network provides examples of wrongful convictions attributed to the misinterpretation of DNA profiles by analysts derived from mixture samples."). *See also* John Butler, Fundamentals at 292 (examining scandals at the FBI and Houston Police Department DNA Units and pointing out that "unfortunately errors in data interpretation by the HPD laboratory led to the false conviction and incarceration of a young man accused of a 1998 rape."); NAS 2009 report at 132 ("Although DNA laboratories are expected to conduct their examinations under stringent quality controlled environments, errors do occasionally occur. They usually involve situations in which interpretational ambiguities occur or in which samples were inappropriately processed and/or contaminated in the laboratory. Errors also can occur when there are limited amounts of DNA, which limits the amount of test information and increases the chance of misinterpretation."); Adam Liptak, *You Think DNA Evidence is Foolproof? Try Again*, N.Y. TIMES, Mar. 16, 2003 (reporting on DNA typing errors committed by the Houston Police Department), https://www.nytimes.com/2003/03/16/weekinreview/the-nation-you-think-dna-evidence-is-foolproof-try-again.html.

[6] *See, e.g.*, Peter Andrey Smith, *When DNA Implicates the Innocent*, SCIENTIFIC AMERICAN, June 1, 2016, https://www.scientificamerican.com/article/when-dna-implicates-the-innocent/ ("In December 2012 a homeless man named Lukis Anderson was charged with the murder of Raveesh Kumra, a Silicon Valley multimillionaire, based on DNA evidence. The charge carried

are documented errors by the very FBI laboratory that conducted the analysis in this case.[7]

Worse yet, intentional falsification occurs in some cases. An analyst may "fak[e] test results to cover up errors arising from cross-contamination of DNA samples and sample mix-ups." Thompson, Tarnish on the Gold Standard, at 12. Indeed, analysts have been "fired for scientific misconduct, and specifically for falsification of test results," from laboratories operated by the FBI, the Chief Medical Examiner in New York City, the U.S. Army, and private enterprises. *Id.* For example, a DNA analyst in the FBI laboratory failed to use proper controls and falsified laboratory reports, *id.*, "render[ing] over two years worth of her STR work scientifically invalid and unsuitable for use in court." UNITED STATES DEPARTMENT OF JUSTICE, OFFICE OF THE INSPECTOR GENERAL, THE FBI DNA LABORATORY: A REVIEW OF PROTOCOL AND PRACTICE VULNERABILITIES (May 2004), available at https://oig.justice.gov/sites/default/files/archive/special/0405/index.htm (last visited Jan. 4, 2021).

---

a possible death sentence. But Anderson was not guilty. He had a rock-solid alibi: drunk and nearly comatose, Anderson had been hospitalized—and under constant medical supervision—the night of the murder in November. Later his legal team learned his DNA made its way to the crime scene by way of the paramedics who had arrived at Kumra's residence. They had treated Anderson earlier on the same day—inadvertently 'planting' the evidence at the crime scene more than three hours later."); Thompson, *Tarnish on the Gold Standard* at 10–11.

[7] *See* Dan E. Krane, *Time for DNA Database Disclosure*, 60 J FORENSIC SCI. 1668 (2015) ("[T]he FBI only recently recognized that dozens of such errors existed in [their population study] data sets after it embarked upon concordance studies with new test kits. . . . The FBI's assertion that the impact of the errors it has 'discovered' on statistical weights that have been generated for tens of thousands of criminal investigations is 'nominal' may be self-serving. Only careful, independent analysis of the corrected underlying genotypes (which have also not yet been made available) will allow that to be determined.").

Understanding why and how these errors occur and why they are of such concern in this particular case requires a brief description of what little is known about the government's DNA evidence and a further explication of the theory of forensic DNA testing.

II. THE DNA EVIDENCE IN THIS CASE

The 16-page DNA Report in this case, Exhibit 1, is a typical FBI laboratory report: short and unilluminating as to methodology or the bases of the expert's opinions. The report states in conclusory fashion that [redacted], with the bulk of the report being devoted to describing the tested items and the bottom-line conclusion reached as to each. Among the items of evidence that were tested for DNA were [redacted]. Unusually, and unlike most other forensic DNA reports viewed by defense counsel, the report does not even list the DNA loci being compared. As for methodology, the report states only that [redacted] *Id.* at 3. A footnote states that [redacted] *Id.* at 14 n.1. In the Remarks section of the report, it states that [redacted][8] *Id.* at 15.

[8] The "FBI file" has not been disclosed to the defense and presumably refers to the case litigation packets being sought in this motion.

The report also describes the serological test performed on various items as follows: [redacted] *Id.* at 3, 6, 13. As to these items, the report states that [redacted] *Id.* at 14–15. The report indicates that [redacted] *Id.* at 14. [redacted] Conclusions based on DNA mixtures allegedly linking individuals to items of evidence in Mr. Bowers' case are far from uncontroversial. The methods and techniques employed by the FBI DNA Laboratory—among them testing low template samples, making copies of DNA fragments using the PCR process which generates known artifacts in the testing results, using STR length-based testing technology rather than DNA sequencing technology, and employing a probabilistic genotyping program like STRmix™—are known to the broader scientific community in many applications to be unreliable and invalid. At this stage of the case, the defense believes that the

[9] It is unclear from the report whether the analyst simply failed to use the confirmatory Takayama Hemochromogen test on the other items or whether the test was used and yielded negative results. Mr. Bowers intends to file a separate motion challenging the government's serology results, but in order to do so he needs the litigation packets to determine exactly what serology methodology the government used on each item of evidence and what the results of the testing showed.

underlying case litigation packets are necessary and essential to determine what specific methodology was used to reach the conclusions stated in the report and to determine through independent expert analysis whether furher grounds in support of this motion exist. These case litigation packets will reveal the techniques used in more detail and provide an opportunity for the defense to present evidence as to the standards that should be employed to reach conclusions or offer opinions purported to have a scientific basis.[10]

The report by the government analyst indicates that she relied on invalid and unreliable techniques and offers unreliable and invalid conclusions regarding the DNA testing. The lack of scientific basis for the testing and conclusions is critical as to two issues: first, the analyst calculated and reports DNA likelihood ratio (LR) statistics, which purport to describe the statistical chance the DNA matches a specific individual in this case versus a random unknown person; and second, the analyst used a technique that is dependent on measuring allelic length and not sequence, and is generated using PCR technology, which is known to run the risk of producing unreliable DNA results.

Multiple firearms were recovered at the scene and were subjected to serological and nuclear deoxyribonucleic acid (DNA) typing using short tandem repeats (STRs). Similar testing was also performed on known buccal samples collected from Mr. Bowers and [redacted]

[10] The defense has previously sought, and this Court has denied, discovery related to laboratory data and case file materials in this case. However, due to the delays in the case proceedings caused by COVID-19, and given that the complex DNA issues involved necessarily require extensive discovery requests, briefing, and hearings, Mr. Bowers, though counsel, believes that now is precisely the time to focus on these pre-trial issues, as was argued on November 2, 2020.

[redacted]

Analyst Plaza [redacted] According to her conclusory report, which does not appear in draft form, she [redacted], which assign statistical probabilities to the DNA results. These probabilities are used to assert [redacted]. The alleged connections are based on a number of assumptions, the presence and absence of certain genetic information, and theories about whether unknown individuals would be expected to better explain the evidence. None of this is intuitively clear to the layperson or counsel, which underscores the critical need to independently review all of the data and information generated in the testing to understand the conclusions that have been drawn by Analyst Plaza through her work and to challenge them before this Court.

The report itself claims that [redacted] Alternatively, some of the evidence appears to be unrelated to

[11] The likelihood ratio is a statistical approach that compares the probabilities of observing the DNA results under two alternative propositions, offering three possible conclusions: exclusion, inclusion, and inconclusive. The LR provides a summary statistic of the probability that the individual, compared to all other people, fits in one of those categories. Calculations are then performed using African American, Caucasian, Southeastern Hispanic and Southwestern Hispanic populations. The lowest calculated likelihood ratio is reported.

anyone to whom it was compared and has an unknown origin. This DNA evidence, outlined further below, as well as the FBI's "Qualitative Equivalencies,"[12] which accompany these statistics, need to be scrutinized closely and should be be excluded because they are inherently unreliable, invalid, and more prejudicial than probative as evidence in this case.

Further, the FBI appears to rely solely on a more antiquated DNA techonology that runs grave risks of producing unreliable results. Recent advancements in DNA sequencing technology provide the opportunity to reassess the reliance on PCR technology and length measurements of particular segments of the genome as the basis for claiming uniqueness in DNA results. This is an analysis of only a fraction of the genetic information available and ignores other distinguishing information, which more advanced technology is now capable of detecting. As discussed below, the reliance on older techniques is unreliable, providing only a veneer of scientific information, which is likely to confuse lay factfinders, instead of conducting straightforward specific genotyping. As a threshold matter, it is also impossible without the case litigation packets

[12] For the FBI Laboratory, likelihood ratio ranges provide the following support for the conclusion:

Likelihood Ratios:	Qualitative Equivalent:
≤1/100	**Exclusion**
>1/100 to ½	Limited Support for **Exclusion**
1	Uninformative
2 to <100	Limited Support for **Inclusion**
100 to <10,000	Moderate Support for **Inclusion**
10,000 to <1,000,000	Strong Support for **Inclusion**
≥1,000,000	Very Strong Support for **Inclusion**

to even know what the actual data results were in this case; for instance, which loci provided readable data and what allelic lengths were actually measured or detected.

As indicated above, the FBI Laboratory, like other DNA laboratories, has a documented history of introducing erroneous and/or falsified DNA evidence in criminal cases.[13] Given this history, the government is required to prove by a preponderance of evidence that the FBI's DNA testing results and statistical analysis are relevant, reliable, and not substantially outweighed by the risk of confusing or misleading the jury. The danger of misleading and confusing the jury with unreliable evidence is especially great with respect to the FBI's use of length-based STR testing and its use of LR analysis of certain mixed samples. Many of the LRs included in the report of Analyst Plaza are fairly categorized as extremely weak and are expected to be LRs found within an uninformative range of data at best, and, at worst, in a range known to produce both false inclusions and false exclusions in internal validations for the use of STRmix.

III. DNA EVIDENCE: OVERVIEW OF THE STATE OF THE SCIENCE

A. How Forensic DNA Typing Works

1) The Theory and Limited Nature of Forensic DNA Testing

[13] Prior to the use of Probabilistic Genotyping software like the STRmix program, the FBI and other DNA laboratories used a statistical approach to complex mixtures referred to commonly as the Combined Probability of Inclusion (CPI) or Combined Probability of Exclusion (CPE). In 1999 and 2001 the FBI STR Population Database contained errors that directly impacted the reliability of the CPI/CPE statistics that were generated in case work. The FBI's own empirical testing to assess the impact of the errors found that as much as a two-fold difference could be erroneously reported out. Other laboratories have confirmed this error in their own case work based on their reliance on the FBI database. Once discovered, most laboratories began to abandon the use of the CPI/CPE.

Deoxyribonucleic acid, or DNA, is a double-stranded molecule that coils to form the characteristic double helix, and is found in all cells possessing a nucleus.[14] Butler, Fundamentals at 19. Forensic DNA typing examines certain locations, or loci, on the DNA strand. The DNA typing technique at issue in this case is short tandem repeat (STR) testing. STR typing measures how many times a short piece of DNA repeats at each of the tested loci; the number of repeats is known as an allele. *Id* at 148. An individual's genetic type, or profile, is the compilation of his or her alleles at each locus tested. At each locus, an individual possesses two alleles: one allele inherited from each biological parent. *Id* at 25. Thus, an individual's DNA profile is simply a list of two numbers per locus examined.[15] An individual can inherit the same allele—i.e., same number of repeats—at a locus from his or her biological parents (*e.g.,* 12, 12). This means the individual is a homozygote at that locus. Alternatively, an individual can inherit two different alleles—two different numbers—at a locus from his or her parents (*e.g.,* 12, 16). This means they are a heterozygote at that location.

With the exception of identical twins, the genome (*i.e.*, the complete genetic composition) of each individual is unique and is inherited from an individual's parents

[14] Most cells, with the exception of red blood cells, possess nuclei. When it is in the nucleus, DNA is tightly packaged into two sets of 23 chromosomes; one set of 23 chromosomes is inherited from each parent. Sperm and egg cells possess only one set of 23 chromosomes each; when they unite, the resulting embryo possesses the full set of 46 chromosomes. Butler, Fundamentals at 23.

[15] The FBI Quantico DNA Laboratory uses the Applied Biosystems' GlobalFiler™ test kit, which tests 23 loci plus amelogenin, an indicator for male or female sex. Thus, an individual DNA profile generated by the FBI lists 46 numbers, or alleles: two for each locus (1 per chromosome) at each of the 23 loci tested, plus either X, X or X, Y at amelogenin. As indicated above, the report in this case does not reveal the results for any of the 23 loci.

with one half coming from the mother and one-half from the father. "However, to limit the expense and time of testing, forensic DNA analysis only examines a small subset of genetic variation within the human genome in order to differentiate among individuals. . . . Out of the more than 6 billion nucleotides present in the diploid human genome, fewer than 4000 nucleotides, or 0.0006% of the material, are examined from highly variable and nondescript regions. By way of comparison, if each nucleotide were 1 inch (2.5 cm) in length, the nuclear DNA in a single cell would be more than 100,000 miles (160,000 km) long—yet forensic DNA tests only examine about 300 feet (90 m) of this information. Because only a fraction of the available DNA information is examined, statistical calculations are performed to estimate the occurrence of a random match based on measured frequencies of particular genetic attributes among unrelated individuals." *Id.*

To further understand the limited nature of the forensic DNA testing conducted in this case it is important to appreciate that in the past twenty years DNA testing techniques have evolved rapidly, from RFLP testing, to DQ Alpha testing, to Polymarker testing, to D1S80 testing, to STR testing using commercially available "kits," and, most recently, to sequence-based DNA testing methods. As indicated above, the only DNA testing at issue in this case is STR testing using the GlobalFiler™ test kit. This STR test kit used in this case employs technology and test procedures substantially different from previous DNA testing methods, and from sequence-based testing.[16]

[16] To greatly simplify an extremely complex subject, the DQ Alpha and Polymarker test kits used for a number of years employed a methodology which determines the *sequence* of a section of DNA by a method called reverse dot blot hybridization. The D1S80 kit, as well as the RFLP method, determined the *length* of a section of DNA by a method called gel electrophoresis. In all

When the first widely-used STR typing kits (Profiler Plus and CoFiler kits) were introduced in 1997–1998, the kits examined only the thirteen STR loci selected by the FBI for CODIS, the national DNA database. In 2001, a new kit called Identifiler was introduced, which examined two additional loci. The even newer GlobalFiler™ test kit used in this case examines twenty-one STR loci, including the original thirteen CODIS loci, as well as one marker on the Y chromosome (Y indel), and Amelogenin (sex determining marker).[17]

---

three of these methods, the lab technician makes a purported match by visually inspecting and comparing either the blue dots produced by the DQ Alpha or Polymarker test kits or the gel typing bands produced in RFLP and D1S80 testing. In contrast, short tandem repeat (STR) testing of DNA fragments using STR kits is fully automated and is done on a sophisticated typing machine (the 310, 3130, or 3500 Genetic Analyzer), which employs a complex proprietary software package (GeneMapper ID-X), and a completely different technology called capillary electrophoresis. The only similarity between STR testing and the older methodology is that like the D1S80 kit, as well as the earliest used RFLP method, STR testing determines only the *length* of a section of DNA, not the sequence of the DNA. The implications of this limitation are explored fully below.

[17] To put the limited nature of a DNA analysis of only thirteen or even twenty-three STR loci into perspective, Dr. Butler notes that "more than 20,000 tetranucleotide STR repeats have been located throughout the human genome. However, when the core STR loci that are widely used today were selected back in the mid-1990s, only a handful of STR loci were known and characterized." Butler, Fundamentals at 148. In Dr. Butler's 2012 book he notes that beginning in March 2007, a "next Generation" of STR typing kits began to be introduced which supposedly addressed such problems as PCR inhibition and included new loci, and that as of 2012 there were "23 core and commonly used loci." Butler, DNA Methodology at 109, 114. In a 2015 article, Dr. Butler elaborates that "[e]xisting DNA databases, now numbering in the millions of STR profiles, make it less likeley that the forensic community will change to different genetic markers in the foreseeable future. Expansion to additional STRs while retaining connection to legacy STR profile information appears to be the way forward in Europe and the United States." J. Butler, et. al*., STR allele sequence variation: Current knowledge and future issues*, 18 FORENSIC SCIENCE INTERNATIONAL: GENETICS 118 (2015). He adds that "[g]iven current information [including information about STR sequence variation], some of [the legacy STR loci] would not be chosen today, but they will likely continue to be genotyped for compatability to large existing DNA databases." *Id.* at 19.

In other words, the FBI may have bought a pig in a poke in selecting the core thirteen loci and building a system that depends on kits like Profiler Plus, CoFiler, and GlobalFiler™ to conduct length-based testing at these loci. The basic problem with this approach is that "analysis of PCR

2) Overview Of The DNA Testing Process

The DNA testing process proceeds via a series of steps: extraction, quantitation, amplification, analysis of genetic data, evaluation, comparison, and interpretation. The first step in generating a DNA profile from a sample is extraction, where the analyst attempts to isolate the DNA and separate it from all other cellular material and debris. Butler, Fundamentals at 99. After extraction of the DNA, the sample is quantitated, *i.e.*, the total amount of DNA present in the sample is estimated. *Id*. at 114. Based on the estimated amount of DNA present, some portion of the extracted DNA is then amplified. Amplification is a process by which DNA is copied at targeted locations (*i.e.*, loci) many times over, generating on the order of a billion copies.[18] *Id*. at 125–26. During the amplification process, the targeted DNA may not amplify if there is an insufficient amount of DNA to start with,[19] or if it is degraded (*i.e*., broken into pieces due to environmental exposure or other stressors), or if there are inhibitors (such as some fabric dyes or excess salts) present in the sample. *Id*. at 68. When targeted DNA does not

---

product length alone fails to capture the potential internal sequence variation that may exist in many STR loci detected via base composition mass spectrometry or through full sequence analysis." *Id.* at 118. As argued below, the failure to account for sequence variation renders the current testing systems no longer reliable enough to use in a criminal case. Further, the constant addition of new loci calls into question the reliability and discriminatory power of now discarded kits such as Profiler Plus and CoFiler. If the astronomical random match probabilities typically generated from thirteen loci matches produced by Profiler Plus and CoFiler were sound, then why would there be a need to constantly test at additional loci?

[18] Amplification is conducted via a technique called polymerase chain reaction, commonly notated as PCR.

[19] Quantitation gives a preliminary estimate of whether the amount of DNA in the extract falls into this low-level range. However, a seemingly sufficient total amount of DNA may be comprised of low levels of DNA from multiple contributors; this is not something that can be discerned from the quantitation step, which does not distinguish between contributors but rather reports the total amount of DNA present.

amplify, that genetic information is lost in downstream steps; this loss of genetic information is known as allelic dropout, a concept discussed further below. *Id*. at 222.

The post-amplification sample consists of large numbers of only the copied alleles, which can then be separated on an instrument called a genetic analyzer so that each allele can be distinguished and then recorded. *Id*. at 175. The result of this process is a series of peaks on a graph, called an electropherogram. *Id*. at 194. The analyst interprets the electropherogram, generating a genetic profile for the evidence sample. Part of the interpretation process involves determining whether the peaks present represent "real" DNA or artifacts of the testing process. Each "real" DNA peak corresponds to an allele present in the sample and the height of each peak roughly corresponds to how much of that allele is present (*i.e*., a taller peak indicates more of a particular allele present). When testing a single source evidence sample (*i.e*., a sample originating from one individual), two peaks of roughly equivalent height should be observed at each locus where the contributing individual is a heterozygote (*i.e.*, possesses 2 different alleles). At loci where the contributor is a homozygote (*i.e*., possesses two of the same allele), one, relatively high peak should be observed, because the two alleles "stack" on top of one another.

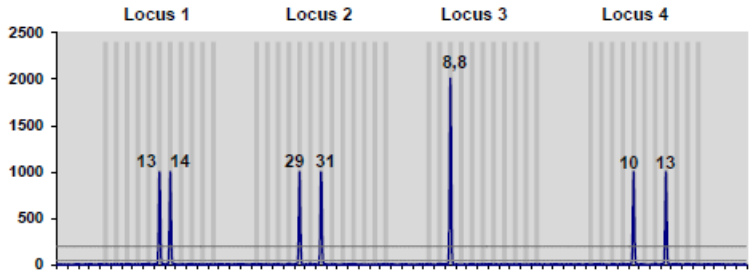

Figure 1. Electropherogram showing ideal, single-source DNA data at four hypothetical loci. Note that at Locus 3, where the DNA contributor is a homozygote (possesses two "8" alleles), his two alleles "stack" on top of one another, resulting in a single peak on the electropherogram. At each of the

other three loci, where the contributor is a heterozygote (i.e. possesses two different alleles), two peaks are observed. Figure from Butler, ADVANCED TOPICS IN FORENSIC DNA TYPING: INTERPRETATION 11, Fig. 1.5 (2014).

The resulting DNA profile for a sample, which is a combination of individual STR genotypes, is compared to other samples. In the case of a forensic investigation, these other samples would include known reference samples such as those taken from the victim or suspects, which are compared to the crime scene evidence. "If there is no match between the questioned forensic sample and the known sample, then the samples may be considered to have originated from different sources." Butler, Fundamentals at 7. Importantly, "[i]n forensic DNA typing, if any STR locus fails to match when comparing the genotypes between two or more samples, then the profiles between the questioned and reference sample will be declared a non-match, regardless of how many other loci match." Butler, Fundamentals at 221 (emphasis added); *see* John Butler, ADVANCED TOPICS IN FORENSIC DNA TYPING: INTERPRETATION 17–18 (2014) ("Butler, Advanced Topics") (emphasis added) ("In forensic DNA Q-K comparisons (as currently practiced in many parts of the world), if *any* STR locus fails to match when comparing the genotypes between two or more samples, then comparison of profiles between the questioned and reference sample is usually declared a non-match, regardless of how many other loci match.").

If the analyst determines that one of the reference profiles "match" or "cannot be excluded from" the evidence profile, the analyst calculates a rarity statistic to contextualize the significance of the match or inclusion. Statistical calculations are the second step in the interpretation process, giving the trier of fact a means of assessing the

possibility that the inclusion is coincidental. "The statistical calculation step is the pivotal element of DNA analysis, for the evidence means nothing without a determination of the statistical significance of a match of DNA patterns." *People v. Barney*, 8 Cal. App. 4th 798, 817 (Cal. Ct. App. 1992). This is because "it would not be scientifically justifiable to speak of a match as proof of identity in the absence of underlying data that permit some reasonable estimate of how rare the matching characteristics actually are." NRC II at 192; *see* NRC I at 74 ("To say that two patterns match, without providing any scientifically valid estimate (or at least, an upper bound) of the frequency with which such matches might occur by chance, is meaningless"). The Scientific Working Group for DNA Analysis Methods, or SWGDAM,[20] created to provide discipline-wide guidelines, similarly admonishes that analysts "must perform statistical analysis in support of any inclusion that is determined to be relevant in the context of a case, irrespective of the number of alleles detected and the quantitative value of the statistical analysis." SWGDAM INTERPRETATION GUIDELINE 4.1 (2010); *see United States v. Cuff*, 37 F. Supp. 2d 279, 282 (S.D.N.Y.1999) ("Without statistical data on the frequency of the matching characteristics in the relevant reference population . . . the jury [is] left to speculate about the value of the DNA evidence.").

3) The Problem of Primer Binding Site Mutations

A disturbing feature of both the Applied Biosystems' STR kits and the competing Promega Corporation kits is that the results that are obtained from one kit can vary from

[20] SWGDAM is an advisory group convened by the Federal Bureau of Investigation. https://www.swgdam.org/ (last visited Jan. 3, 2021).

the results obtained using the competitor's kit, or even a kit from the same manufacturer that uses a differant set of primers (a phenomenon called "null alleles" in the literature).[21] For example, in May 2020, it was reported to NIST by the Iowa State Crime Lab that a sample tested at D12S391 as a "22, 22" using the GlobalFiler kit, but tested at the same locus as a "21, 22" using Promega's PowerPlex Fusion 5C kit. Similarly, the Florida Department of Law Enforcement recently reported to NIST that a sample tested at FGA as a "19.2, 25" using the GlobalFiler kit, but tested at the same locus as a "19.2, 19.2" using Applied Biosystems's Identifiler Plus kit. If the Court follows the reasoning of cases such as *United States v. Trala,* 162 F. Supp. 2d 336, 348 (D. Del. 2001), then both results and both kits would be considered reliable since both use PCR and STR testing. This is logically inconsistent and would lead to absurd results. For example, if one lab tested the evidence using a GlobalFiler kit and the results were inculpatory, and then

[21] When null alleles are discovered, they are catalogued on the Short Tandem Repeat DNA Internet DataBase (STRBase) maintained by the National Institute of Standards and Technology and available at https://strbase.nist.gov/NullAlleles.htm (last visited Jan. 3, 2021). The NIST website explains that "[t]he presence of sequence variation in the flanking region of commonly used STR loci along with the use of different PCR primers for typing these STR repeats can result in failure to amplify one or both of the alleles present at the particular locus. These 'null' or 'silent' alleles due to allele dropout are typically surmised with concordance studies where more than one PCR primer set is used to test the STR locus of interest." As of December 31, 2020, the NIST website reported the discovery of null alleles for the following STR loci: D12S391,CSF1PO, FGA, THO1, TPOX, VWA, D57818, D7S820, D8S1179, D13S317, D16S659, D18S51, D21S11, D19S433, DYS533, and Amelogenin. These 16 loci are all tested for in the GlobalFiler kit used in this case. According to the NIST website, these discoveries were made between 1998 and October 2, 2020. In fact, discoveries of null alleles continue to be made and reported in the scientific literature, although not cataloged on the NIST websites. *See, e.g.*, Y. Yao, et. al, *Null alleles and sequence variations at primer binding sites of STR loci within multiplex typing systems*, 30 LEGAL MEDICINE 10 (2018) ("Yao, Null Alleles") (reporting the discovery of new null alleles typed with the Powerplex 21 kit for THO1, FGA, D5S818, D8S1179, and D16S539). Further, as in other studies, the discovery of null alleles was made by Yao not from concordance studies, but from analysis of expected patterns in paternity cases.

another lab retested the same evidence with Promega's PowerPlex Fusion 5C kit, or with Applied Biosystems's Identifiler Plus kit, and the results were exculpatory, then under *Trala* a court would have to conclude that both kits were reliable and generally accepted because they both used PCR and both involved STR testing. This example illustrates the fallacy of the argument and the reasoning that STR-based tests are interchangeable because they involve STR markers as a class or PCR as a common element.

The inconsistency in test results using different kits or primers was first documented in a series of articles that addressed serious problems concerning the inability of a particular STR testing kit to produce results consistent with other kits or other primers. These articles also demonstrated that the ability to investigate this problem was and is being impeded by Applied Biosystem's refusal to disclose their primer sequences.[22] *See, eg.*, David Parsons, *Resolution of a Typing Differance Between Perkin*

[22] More recent studies indicate that Promega's refusal to reveal their primer sequences is also impeding investigation of null alleles. *See, e.g.*, Yao, Null Alleles at 11 (emphasis added) ("Although the primer sequence for STR loci of PP21 is not available, the approximate primer binding site could be inferred based on previous reports. Therefore, a single base variation at inferred primer binding site *should* be responsible for the dropout of alleles in STR genotyping."). Yao notes that "concerns have been raised about null alleles since a genotyping error from commercial STR typing systems or a discrepancy between them could bring up a false or inaccurate interpretation in forensic practice. Therefore, it is important to characterize variants at primer binding sites of STRs within multiplex typing systems to eliminate genotyping error or discrepancy in individual identification and paternity testing." *Id.* at 10. Yao concludes:

> In the establishment and utilization of forensic DNA databases, typing discrepancies generated by using different STR typing kits because of null alleles can impair the identification of individuals, familial searching, and data exchange. Therefore, more attention on STR loci with null alleles should be paid in forensic casework, especially on those more frequently observed ones. Furthermore, sequence variations at primer binding sites should be characterized, which might help in the development of more efficient primer sets for STR typing.

Elmer's Profiler Plus Kit and Promega's PowerPlex 1.1 Kit Using Sequence Analysis, TENTH INTERNATIONAL SYMPOSIUM ON HUMAN IDENTIFICATION (1999) (Two samples typed as a homozygous at vWA locus using Profiler Plus kit, but as heterozygous using Promega's PowerPlex kit. A primer binding site mutation was suspected as the cause, but "(s)ince the sequence for the primers for the Profiler Plus are not published, the exact location of the mutation with respect to the primer cannot be known. If this mutation occurred in the priming region for Perkin Elmer's Profiler Plus kit, it would provide the needed explanation for the observed results.");[23] C. Alves, et. al., *vWA STR Genotyping: Inconsistency Between Perkin Elmer's Profiler Plus Kit And Promega's Geneprint*, INTERNATIONAL SOCIETY FOR FORENSIC HAEMOGENITICS, EIGHTEENTH INTERNATIONAL CONGRESS ABSTRACTS, August 17–21, 1999, San Francisco, California, at 30 (simultaneous study of vWA locus by the Applied Biosystems' Profiler Plus Kit and the Promega Geneprint Kit produced an inconsistency between the genotyping in each kit: using Profiler Plus it was found to be "18, 18" and with Geneprint "16, 18." "Since primer sequences were not available from the manufacturers we could not sequence the corresponding regions. However, it is tempting to interpret the inconsistency as a result of a Perkin Elmer primer annealing failure . . . . The finding now reported evidences the need for caution when comparing genotypes or gene frequencies made in amplicons and

---

Id. at 13.

Dr. Butler claims in his 2010 book that, unlike Applied Biosystems, it is Promega's policy to reveal its primer sequences. Butler, Fundamentals at 162. Apparently, the policy has changed.

[23] https://promega.media/-/media/files/resources/conference-proceedings/ishi-10/poster-abstracts/70parsons.pdf?la=en (last visited Jan. 3, 2021).

produced by different primers."); M.C. Kline, et. al., *Nonamplification of a vWA Allele*, J. FORENSIC SCI. 1998 Jan., 43(1):250 (National Institute of Standards and Technology researcher documents same inconsistency and indicates that Perkin Elmer (predessor to Applied Biosystems) "is aware of the problem and they are actively pursuing an explanation for this allelic dropout by sequencing the sample"); S. Walsh, *Commentary on Kline, MC, Non-Amplification of a vWA Allele*, J. FORENSIC SCI. 1998 Sept., 43(5) 1103 (Perkin Elmer admits the problem exists, and claims it is caused by a flanking sequence mutation. "Our laboratory has observed flanking sequence mutations in several STR loci, including the vWA loci reported here, D16S539, and TPOX. Other laboratories have reported flanking sequence mutations at D13S317 and DS7820." Perkin Elmer admits that the problem will continue, but claims that it can be avoided by using Perkin Elmer products exclusively.).

The NIST website indicates, and the 2018 Yao article confirms, that "null alleles at most CODIS STR loci have been observed to date." Yao, Null Alleles at 11. The Yao article also notes that "[s]ince it is difficult to pick out every dropout of alleles at STR loci caused by primer binding site mutations, null alleles can be frequently overlooked in STR typing. Furthermore, null alleles might be overlooked or be mistaken as mutation events in paternity testing. Therefore, the number of null alleles observed in previous reports should be underestimated." *Id.* at 12. Dr. Butler seems to agree, stressing in his 2015 book that "[i]t is important to emphasize that *no primer set is completely immune to the phenomenon of null alleles*." Butler, Advanced Topics at 102 (emphasis in original). *See* S. Inokuchi, *Identification of a common single nucleotide polymorphism at the*

primer binding site of D2S1360 that causes heterozygote peak imbalance when using the Investigator HDplex Kit, 131 INT. J. LEGAL MED. 1531 (2017) (finding that 22 of 196 sampled Japanese individuals had a primer binding site mutation at D2S130 and noting that "[t]he estimated allele frequency of the mutation is appreciably higher than the allele frequency of the null allele at D19S433 in the core STR locus in the Japanese population"; also reporting that in their attempts to study the problem the authors were told by the manufacturer of the kit that "the primer sequence of D2S1360 is not to be disclosed."); M. Kline, *STR sequence analysis for characterizing normal, variant, and null alleles*, 5 FORENSIC SCIENCE INTERNATIONAL: GENETICS 329 (2011) 329–332 ("DNA sequence variation is known to exist in the repeat region and the flanking regions of most of the widely-used STRs. A mismatch due to sequence variation in the DNA template can cause allele dropout when it falls within PCR primer binding sites."). Based on these recent studies, it is reasonable to infer that the null allele problem first observed in relation to VWA is not confined to that locus and that in this case it extends to any loci where either Mr. Bowers' reference sample or any comparison sample was typed as a homozygote.[24]

It is also important to stress that the problem is not confined to Applied Biosystems kits but extends to Promega and other kits as well. The NIST website cites

[24] In this case, because the defense has so far been denied access to the litigation packets and the disclosed DNA report does not indicate the DNA typing results for the 23 loci tested in the GlobalFiler kit for any of the samples, it is not known whether either Mr. Bowers or the comparison samples tested as a homozygote at any of the 23 loci. This information, which will obviously be included in the litigation packets, is critical to determining whether null alleles present a problem in this case.

several examples where an allele failed to detect when the Powerplex 16 kit was used, and several other instances where an allele failed to be detected when a GlobalFiler or Profiler Plus or other Applied Biosystem kit or other kit was used. The 2018 Yao article indicates that the problem persists even with the latest and greatest of Promega's kits, the Powerplex 21 kit, which presumably includes Promega's best efforts to address the problem. As Yao notes, "[e]fforts have been made to obtain ideal primers, which optimize the assay to promote both specificity and efficiency of PCR amplification. However, rare variants at primer binding sites inevitably occur among populations ." Yao, Null Alleles at 10.

The Yao article also addresses the issue of whether the problem can be solved by simply using two separate kits with two separate sets of primers to test the same sample. The article concludes:

> Since commercial amplification kits usually use different primer sets for the same locus, genotypes of STRs with null alleles could be possibly corrected by using other multiplex kits. In this study, genotypes of the 5 cases with null alleles obtained using PP21 were corrected using [two other kits, EX22 and Identifiler Plus]. However, it should be noted that not a single STR typing system could avoid all potential null alleles since rare variants occur at random. In fact, cases with null alleles have been reported when genotyping using EX22, or Identifiler Plus kit and AmpF.STRR IdentifilerR PCR Amplification Kit (Identifiler), which employ the same loci and primer sets. Rare circumstances additionally happened under which two or more STR loci with null alleles were observed in a single case detected by the same commercial multiplex kit.

Id. at 11.

Of course, since neither Applied Biosystems nor Promega will release their primer sequences, there is no way to verify whether different kits are in fact using different primers, and, if so, whether the primers are the root of the problem.

In any event, it is noteworthy that one of the studies on this issue is A. Gavrilidis, et. al., *Mutations, Variations and Missing Alleles in the STR Systems*, TENTH ANNUAL SYMPOSIUM ON HUMAN IDENTIFICATION (1999).[25] This study indicates that "[w]e report the failure to amplify and detect fragments at both the D135317 and Amelogenin loci. Both the Promega PowerPlex™ and PE Applied Biosystems AmpF/STR™ Profiler Plus kits failed to amplify in both instances. This may indicate sequence variation at these primer positions." *Id*. Dr Butler's 2015 book reports on a study that found that when a sample was typed with a Powerplex ESX 17 kit and then typed with an ESS plex SE kit, the concordant result was "13, 15." However, when the sample was typed with a NGM Select kit, the result was "15, 15." Butler, Advanced Topics at 100. In other words, for the reasons explained in the Yao article, the problem is not solved by testing with two different kits from two different manufacturers.

A number of other remedies have been suggested to control for this problem. In Dr. Butler's book, "Forensic DNA Typing", he first offered the solution that

> when identical primer sets are used to amplify evidence samples and suspect reference samples, full concordance is expected *from biological materials originating from a common source*. If the DNA templates and PCR conditions are identical between two samples *from the same individual*, then identical DNA profiles should result regardless of how well or poorly the PCR primers amplify the DNA template. The potential of null alleles is not

[25] https://promega.media/-/media/files/resources/conference-proceedings/ishi-10/poster-abstracts/02gavrilidis.pdf?la=en (last visited Jan. 3, 2021).

> a problem within a laboratory that uses the same primer set to amplify a particular STR marker.

John M. Butler, FORENSIC DNA TYPING 90–93 (2001) ("Butler, Forensic DNA Typing"); Butler, Fundamentals at 102 (making the same statement).

The problem with this logic is that the theory propounded by Dr. Butler appears to be correct as far as it goes, which is only as far as a guilty defendant. But in cases where the defendant is *not* the perpetrator, the evidence sample from the perpetrator and the reference sample from the defendant originate from different people (because the defendant did not commit the crime and did not leave his DNA at the crime scene) and thus the samples contain different DNA. One sample may contain DNA with a mutation and the other sample may not. Allelic dropout may occur in one sample and not the other. Under these circumstances, allelic dropout in one sample may lead to a false match that falsely incriminates an innocent defendant.

Thus, allelic dropout is not always a benign phenomenon in criminal cases because not all defendants are guilty. Where the perpetrator and the defendant are different people and where one of them is mutant heterozygous and the other is normal homozygous, the dropout in the mutant heterozygous sample can cause a false homozygosity and a false match to the homozygous sample. This means, in theory at least, that allelic dropout is capable of leading to the conviction of innocent defendants. If this conclusions is accurate, then Butler's idea that allelic dropout cannot cause false results in a criminal case as long as the same primers/kit are used on both the defendant's

and the perpetrator's DNA samples is a very serious falsehood based on the improper assumption that the defendant is guilty.

Dr. Butler also offered other solutions. "First, the problem PCR primer could be redesigned and moved away from the problematic site. . . . However, this solution could result in the new primer interfering with another one in the multiplex set of primers or call for new PCR reaction optimization experiments. Clearly this solution is undesirable because it is time consuming and labor intensive." Butler, Forensic DNA Typing at 32.

"A second solution is to simply drop the STR locus from the multiplex mix rather than attempting to redesign the PCR primers to avoid the site. This approach is only desirable when early in the development cycle of a multiplex STR assay." *Id.* at 33.

"A third, and more favorable, solution is to add a 'degenerate' primer that contains the known sequence polymorphism. This extra primer will then amplify alleles containing the problematic primer binding site sequence variant. However, if the sequence variation at the primer binding site is extremely rare, it may not be worth the effort to add an additional primer to the multiplex primer mix." *Id.* But, as Dr. Butler was quick to add, "[i]t is truly a challenge to design multiplex STR primer sets in which primer binding sites are located in sequence regions that are as highly conserved as possible and yet do not interfere with primers amplifying other loci." *Id.* at 34; *see* Butler, Advanced Topics at 102 (repeating the same statement). And as the Yao article noted in 2018, "[e]fforts have been made to obtain ideal primers, which optimize the assay to promote both specificity and efficiency of PCR amplification. However, rare variants at primer binding sites inevitably occur among populations ." Yao, Null Alleles at 10; *see*

K. Lazaruk, *Sequence variation in humans and other primates at six short tandem repeat loci used in forensic identity testing*, 119 FORENSIC SCIENCE INTERNATIONAL 1, 2 (2001) (article by Applied Biosystems' employees explaining efforts to fix Profiler Plus kit: "The discovery of primer binding site mutations in a number of the STR loci (vWA reverse, vWA forward, TPOX) has led to the conclusion that it is difficult to find an absolutely conserved primer binding region that will not be prone to mutation in some individuals.").

"A fourth possible solution to correct for allele dropout that will work for some problematic primer binding sites is to re-amplify the sample with a lower annealing temperature and thereby reduce the stringency of the primer annealing. If the primer is only slightly destabilized, as detected by a peak height imbalance with a heterozygous sample (Figure 6.8), then it may be possible to correct the peak height imbalance by lowering the annealing temperature during PCR." Butler, Forensic DNA Typing at 34. This approach, however, has been tried and failed to correct the problem. *See* E.M. Dauber, et. al., *Two examples of null alleles at the D19S433 locus due to the same 4 bp deletion in the presumptive primer binding site of the AmpFlSTR Identifiler kit*, FORENSIC SCIENCE INTERNATIONAL: GENETICS SUPPLEMENT SERIES 1 107–108 (2008) ("Lowering the annealing temperature is not a helpful strategy to recover null alleles in case of deletions or insertions in the primer binding site. Merely the use of alternative primers can overcome allelic drop-out in these cases. As primer sequences are not available for the AmpFlSTR1 IdentifilerTM PCR Amplification Kit,

only the manufacturer could theoretically add a degenerate primer, which also amplifies the alleles carrying this deletion.").

Dr. Butler's 2015 *Advanced Topics* book contains, with one important exception, an almost verbatim discussion of the solutions to the problem of null alleles that he offered in his 2005 and 2010 books, including the exact same language that is quoted above. The one exception is this sentence at the very end of the discussion of null alleles: "Perhaps future DNA sequencing techniques will enable a more complete characterization of STR repeats and their flanking regions." Butler, Advanced Topics at 104.

Significantly, Dr. Butler includes in the bibliograhy to this chapter a cite to a 2011 article by Margeret Kline, the NIST researcher who was one of the first to discover null alleles at the vWa loci. That article, M. Kline, *STR sequence analysis for characterizing normal, variant, and null alleles,* 5 FORENSIC SCIENCE INTERNATIONAL: GENETICS 329 329–332 (2011), states unequivocally that "[i]n order to understand at the molecular level the reason for a variant or null allele, the DNA sample needs to be sequenced at the STR locus." *Id.* at 329.[26] That article also informed the forensic science community as follows:

> Our group at NIST has provided variant and null allele sequence analysis as a service to the forensic DNA community *for the past decade* in addition to cataloging this information on the STRBase website.

Id. at 329 (emphasis added).

Finally, the conclusion to the article states:

[26] https://www.nist.gov/system/files/documents/mml/bmd/genetics/STR-Sequence-Analyisis-for-characterizing-normal-variant-and-null-alleles-2011NIST.pdf (last visited Jan. 3, 2021).

> DNA sequence analysis is helpful to understand the molecular basis for an allele that sizes "off-ladder" or one that fails to be amplified with a specific set of PCR primers. *Through funding from the National Institute of Justice, NIST has provided STR allele sequence analysis free-of-charge to the human identity testing community.* The supplemental files with this article provide the DNA sequencing primers used and a selection of results obtained with normal, variant, and null alleles.

Id. at 331 (emphasis added).

This article would seem to put the lie to any suggestion the government may make that sequence-based DNA testing is unavailable to the FBI or, if available, is too costly or impractical to perform in this capital case. As will now be demonstrated, that argument, whatever its merits in years past, has now been completely undermined by recent developments.

4) The Issue of Length versus Sequence Testing

As indicated above, the GlobalFiler DNA kit types the length of a short tandem repeat, not its actual sequence. Sequencing looks at the actual sequence of the DNA building blocks A, C, T, and G, and how they align across a strand of DNA. Importantly, sequence analysis has shown that alleles measuring the same length can have different sequences, which makes them two different alleles.

As Dr. Butler explains in his 2015 Advanced Topics book, it is being discovered with increasing frequency that certain STRs have the exact same length but in fact two different DNA sequences. As he states:

> Complex repeat sequences, such as those found in D21S11, can contain variable repeat blocks in which the order is switched around for alleles that are the same length. For example, the STR locus D21S11 has four alleles that are all 210bp when amplified with the Identifiler kit (Appendix I). While these alleles would be sized based on overall length to be "allele 30," they

> contain repeat blocks of 4-6-CR-12, 5-6-CR-11, 6-5-CR-11, and 6-6-CR-10 for the pattern [TCTA]-[TCTG]-constant region (CR)-[TCTA]. *In such cases,variant alleles would only be detectable with complete sequence analysis*.

Butler, Advanced Topics at 129 (emphasis added).

Dr. Butler's "solution" to this fundamental problem is startling. He states:

> It is important to realize that from an operational point of view internal allele variation is not significant. In the end a match is being made against many loci not just one, such as D21S11, with possible internal sequence variation. Most of the STR loci used in human identity testing have not exhibited internal sequence variation, particularly the simple repeat loci TPOX, CSF1PO, D5S818, D16S539, TH01, D18S51, and D7S820.[27] Alleles are binned based on measured size with STR typing since *sequence analysis of individual alleles is too time consuming* and, because STR variation is primarily size-based, would *rarely* reveal additional information.

Butler, Advanced Topics at 128 (emphasis added).

In other words, simply ignore the problem because the sequencing solution would be "too time consuming," and sequence variations are relatively rare, confined mainly to the D21S11 locus. That cannot be reliable science in a case in which the analyst is

[27] It is interesting to note that in the 2005 edition of his book this sentence read: "Most of the STR loci used in human identity testing have not exhibited internal sequence variation, particularly the simple repeat loci D8S1179, D3S1358, D21S1138, VWA." John Butler, FORENSIC DNA TYPING: BIOLOGY, TECHNOLOGY, AND GENETICS OF STR MARKERS 131–32 (2d ed. 2005) ("Butler, Forensic DNA Typing"). In the 2012 DNA Methodology book, the sentence was changed to read: "Most of the STR loci used in human identity testing have not exhibited internal sequence variation, particularly the simple repeat loci TPOX, CSF1PO, D5S818, D16S539, TH01, D18S51, and D7S820." The sentence had to be changed because between 2005–2012, it was discovered that alleles at the D8S1179, D3S1358, D21S1138, and VWA loci did exhibit internal sequence variation. The situation is a good illustration of how fluid our state of knowledge is about STRs, and how assertions being made by DNA experts rapidly become incorrect as new information is learned. As Butler acknowledged in his 2012 book, "we have listed the reference where each new allele (and its sequence if published) has been described. As more and more samples are analyzed using these STR loci, we recognize that new (rare) alleles will be discovered and that this list will quickly become outdated." Butler, Forensic DNA Typing at 561–62.

prepared to testify that [redacted] (DNA Report, attached as Exh. 1 at 15); most labs, including the FBI, have long had the benefit of free sequencing analysis by NIST; and sequence variation at even one locus remains significant in light of Dr. Butler's own admonition that "[i]n forensic DNA Q-K comparisons (as currently practiced in many parts of the world), if any STR locus fails to match when comparing the genotypes between two or more samples, then the comparison of profiles between the questioned and reference sample is usually declared a non-match, regardless of how many other loci match." Butler, Advanced Topics at 20–21. The FBI is in accord. *See* DNAUI STR Protocol Manual (Oct. 2, 2006) at 28 ("A forensic exclusion is declared when upon comparison of the DNA profile from a reference specimen (e.g., K specimen or Q specimen of established origin) to the DNA profile from a single-source Q specimen, the profiles are found to be different at one or more loci.").

Morevover, contrary to Butler's assertions in his 2015 Advanced Topics book, DNA sequence variations at STR loci are neither rare nor confined to the D21S11 locus, as Butler's own publications and the publications of others demonstrate. Butler first began writing about this topic in his 2001 book, where he briefly stated:

> Complex repeat sequences, such as those found in D21S11, can contain variable repeat blocks in which the order is switched around for alleles that are the same length. For example, the STR locus D21Sl 1 has four alleles that are all 210 bp when amplified with the Profiler Plus TM kit (Appendix I). While these alleles would be sized based on length to be 'allele 30', they contain repeat blocks of 4-6-CR-12, 5-6-CR-ll, 6-5-CR-ll, and 6-6-CR-10 for the pattern [TCTA]-[TCTG]-constant region (CR)-[TCTA]. In such cases, variant alleles would only be detectable with complete sequence analysis.

Butler, Forensic DNA Typing at 88.

Even in 2001, focusing solely on D21S11 was somewhat misleading because in the referenced Appendix, Dr. Butler documented that by 2001 researchers had found several other loci where it had been discovered that there was one or more alleles where individuals had exhibited alleles of the same length but different sequence: FGA; vWA; D3S1358; and D13S317. *Id.* at 274–283. Thus, 5 of the 13 core loci used in forensics had exhibited this phenomenon by 2001. Further, Dr. Butler's 2001 book seemed to indicate that the quest to find such variations was in an early stage: "we have listed the reference where each new allele (and its sequence, if published) has been described. As more and more samples are analyzed using these STR loci, *we recognize that new (rare) alleles will be discovered and that this listing will quickly become outdated.* We encourage the reader to consult the STRBase variant allele listing (http://www.cstl.nist.gov/biotech/strbase/var_tab.htm) and to contribute newly discovered alleles so that they may be categorized for fellow workers in this field." *Id.* at 273–74 (emphasis added).

Consistent with this passage, another scientist had written in 2001 that "[u]nderlying the length differences at an STR locus, some sequence variation is expected simply due to the frequency of sequence variation in human populations. For example, single nucleotide polymorphisms (SNPs) are estimated to occurr in the human genome at a frequency of approximately 1 in 500–1000 nucleotides. When other types of sequence variations, larger insertions or deletions, etc., are included, the frequency of sequence variation is estimated to be 1 in 250–300 nucleotides." K. Lazaruk, *Sequence*

variation in humans and other primates at six short tandem repeat loci used in forensic ID. Entity Testing, 119 FORENSIC SCIENCE INTERNATIONAL 1, 2 (2001).

In his 2005 textbook, Dr. Butler made the identical statement he made in his 2001 textbook, but he now added, without mentioning the Lazaruk article, as follows:

> It is important to realize that from an operational point of view internal allele variation is not significant. In the end a match is being made against many loci not just one, such as D21S11, with possible internal sequence variation. Most of the STR loci used in human identity testing have not exhibited internal sequence variation (see Appendix I), particularly the simple repeat loci TPOX, CSFlPO, D5S818, Dl6S539, THOl, Dl8S51, and D7S820. Remember that we are essentially binning alleles based on measured size anyway with STR typing since sequence analysis of individual alleles is too time consuming and would rarely reveal additional information because STR variation is primarily size-based.

Butler, Forensic DNA Typing at 131–32.

Here, again, the Appendix listed the same five additional loci as listed in the 2001 edition as having one or more alleles that were the same length but different sequence. *Id.* at 562–573. The implication that this issue thus concerned only D21Sll was incorrect.

Dr. Butler did not have a comparable section of a chapter in his 2009 Fundamentals textbook devoted to this topic and there was no Appendix listing the alleles. However, in a chapter on Future Trends, he commented that "[o]ver a decade ago, STR markers were successfully analyzed via MALDI time-of-flight mass spectrometry" and "[i]n this manner, STR alleles that are apparently homozygous by electrophoretic techniques have been subdivided into separate alleles if internal sequence polymorphisms exist." Butler, Fundamentals at 428. He further stated that "[t]he expense of mass spectrometers and expertise required to keep them running and the previous widescale

acceptance of fluorescent methodologies will likely keep mass spectrometry from becoming a major player in forensic DNA analysis of STR markers." *Id.* at 429. He also commented that "[i]t is unclear whether or not next-generation sequencing techniques will help traditional forensic testing as they are helping ancient DNA sample sequencing. Current methods have a difficult time with repetitive sequences and thus unless future improvements are made, STR regions would probably not be able to be reliably analyzed with next-generation DNA sequencing." *Id.* at 428.

In his 2012 textbook, Dr. Butler included a section called "Same Length but Different Sequence" in which he made the exact same statements he made in his 2005 book on this topic, except that he added:

> Over the past several years, work with mass spectometry has demonstrated that same-size STR alleles with different internal sequences or alleles with sequence variation in the flanking region can be resolved from one another using a base composition approach (Oberacher et al. 2008, Pitterl et al. 2008, Planz et al 2009, Pittetl et al. 2010). This sequence variation is sometimes refered to as a SNPSTR, where single nudeotide variation is coupled with STR allele identification. As seen in Appendix 1 certain STR loci, such as SE33, are more prone to internal sequence variation.[28]
>
> A comparison of the variation observed in 11 STRs—the 10 SGM Plus loci and 5E33—in DNA samples from 94 Yakut and 108 Khoisan individuals was performed using conventional STR analysis, mass spectrometric STR analysis, and direct sequencing (Pitterl et al. 2010). The mass spectrometry approach that enabled both sequence and length variation to be detected clearly expanded the number of detectable alleles.

Butler, DNA Methodology at 129. The Appendix now listed four additional loci as having one or more alleles measuring the same length but different sequence: D871179,

[28] SE33 is among the locus tested in the GlobalFiler kit.

D2S1338, D1S1656, and D12S391. *Id.* One of those loci, D871179, was found by Margaret Kline at NIST in 2010 to have three allele 13's, all measuring 147 bp in length, but all having different sequences: 13(a), 13(b), and 13(c). In addition, the 2012 text listed for the first time that at the D3S1358 locus, it had been discovered that there were three allele 15's at this locus, not two, as had been listed in the 2001 book.

In his 2015 textbook, Dr. Butler renamed these types of allele "Isoalleles" and, whereas in his 2012 book he had written that "[a]lleles are binned based on measured size with STR typing since sequence analysis of individual alleles is too time consuming and, because STR variation is primarily size-based, would rarely reveal additional information," he now wrote:

> Sequence analysis of individual alleles has traditonally been too time consuming and would, in most cases, rarely reveal additional information because STR variation is primarily size-based. However, recent work with base composition analysis by mass spectrometry and next-generation sequencing (NGS) has expanded the ability to more readily assess internal variation in STR allele sequence.

Butler, Advanced Topics at 67.

He added:

> STR allele sequencing with a conventional Sanger sequencing approach (e.g. Kline et al. 2011) works well, but is not something that is routinely used since it is slow and labor-intensive. So-called NGS approaches utilize millions of parallel reactions to rapidly generate DNA sequences that can be aligned bioinformatically to a reference sequence. In one study involving D21S11, an NGS approach was used to demonstrate that four apparent homozygotes with CE methods were in fact heterozygous due to different sequence isoalleles (Rockenbauer et al. 2014). Likewise, many novel D12S391 variants were discovered including three 20.3 isoalleles when these alleles were fully sequenced (Dalsgaard et al. 2014).

Id.

As indicated in n.17, *supra*, in 2015 Dr. Butler and other researchers published an article taking the position that:

> STR analysis has traditionally been performed by size-based DNA separations using gel electrophoresis or capillary electrophoresis (CE). However, analysis of PCR product length alone fails to capture the potential internal sequence variation that may exist in many STR loci detected via base composition mass spectrometry or through full sequence analysis. In forensic applications, STR sequence data is expected to increase the effective number of alleles, which will improve discrimination and may aid mixture interpretation in some cases.

Butler, et. al., *STR allele sequence variation: Current knowledge and future issues*, 18 FORENSIC SCIENCE INTERNATIONAL: GENETICS 118 (2015).

According to Dr. Butler and his co-authors,

> Sequence information provides a more in-depth evaluation of STR alleles and three categories of results: (1) loci containing numerous isoalleles where internal sequence variation within the repeat region can enable better resolution when allele sequence information is available, (2) loci that are expected to contain significant flanking region variation within the measured PCR product, and (3) loci with little-to-no further information available beyond the variation that can be detected with electrophoretic sized-based measurements.

Id. at 125.

The radical shift in Dr. Butler's thinking on this issue illustrates the truth of his own observation that "[w]e live in an age of rapid discovery in biotechnology, and new technologies and instruments are continually being developed." Butler, Fundamentals at 427. And, of course, Dr. Butler is not the only one writing about this issue. Since the time Dr. Butler first started writing about this issue in 2001 there has been an explosion of scientific literature extolling the benefits of Next Generation Sequencing and the

limitations of "traditional" STR length comparisons. A representative but by no means exhaustive summary of this literature is as follows:

a. Sarah L. Fordyce, et. al, *High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform*, BIOTECHNIQUES, Vol. 51, No. 2, August 2011, at 127–133 ("Current methods to investigate STR loci, including PCR-based standard fragment analyses and capillary electrophoresis, only provide amplicon lengths that are used to estimate the number of STR repeat units. These methods do not allow for the full resolution of STR base composition that sequencing approaches could provide.");

b. Christophe Van Neste, et. al., *Forensic STR analysis using massive parallel sequencing*, FORENSIC SCIENCE INTERNATIONAL: GENETICS 6 810–818 (2012) ("Second generation sequencing (SGS) . . . can generate individual sequences of the alleles present in an STR amplicon mixture. This way, alleles with the same length can be distinguished based on SNPs or different repeat structures.");

c. Kristiaan J. van der Gaag and Peter de Knijff, *Forensic nomenclature for short tandem repeats updated for sequencing*, FORENSIC SCIENCE INTERNATIONAL: GENETICS SUPPLEMENT SERIES 5 e542–e544 (2015) ("For over two decades, the analysis of short tandem repeats (STRs) in forensics was routinely performed using capillary electrophoresis (CE). With CE, the length of a DNA fragment containing an STR is determined. STR alleles are identified by comparing unknown fragment lengths with a reference allelic ladder containing fragments with known repeat-lengths. The use of a simple number, representing the number of repeats was sufficient as nomenclature for

STR allele variation. Recent developments in massively parallel sequencing (MPS) technologies enable high-throughput sequencing of STRs, revealing additional sequence-variation in many of the STRs.");

d. Xueying Zhaoa, et. al, *Massively parallel sequencing of 10 autosomal STRs in Chinese using the ion torrent personal genome machine (PGM)*, FORENSIC SCIENCE INTERNATIONAL: GENETICS 25 34–38 (2016) ("STR analysis has traditionally been accomplished by length-based separation method using capillary electrophoresis (CE). However, the length-based STR typing method fails to capture the potential sequence variants which may reside within repeat motifs or flanking regions of STR loci. . . . In forensic applications, it is expected that identification of intra-allelic sequence variations may improve the discrimination power and show benefits in characterizing mutational events of STRs, what is more, provide opportunities for better mixture interpretation in some cases. Massively parallel sequencing (MPS) technologies, also termed next generation sequencing, allow parallel sequencing analyses of many thousands of genomic regions in a single reaction.");

e. Fei Guoa, et. al., *Evaluation of the Early Access STR Kit v1 on the Ion Torrent PGMTM platform*, FORENSIC SCIENCE INTERNATIONAL: GENETICS 23 111–120 (2016) ("Although size-based separation by capillary electrophoresis (CE) technology has been the standard method for STR typing, it can only identify STR by length but not by sequence. That is to say isoalleles that alleles are the same in length but different from sequence are unable to be resolved using CE technology. . . . So far, STR

typing based on next generation sequencing (NGS) has the potential not only to reveal the length and sequence variations of STRs but also to meet the demands of high throughput, high multiplexing and high sensitivity. Besides, NGS has no limit on the similar length of fragments between markers so that STR amplicons can be designed as short as possible, which will be beneficial to detect degraded samples [8,9]. Nowadays, more and more researchers are focusing on NGS-STR typing in forensic practice .The year before last, Thermo Fisher Scientific released the first integrated NGS-STR solution, the Ion TorrentTM HID STR 10-plex...In 2015, it has been upgraded to a 25-plex panel 'the Early Access STR Kit v1' including 16 of 20 expanded Combined DNA Index System (CODIS) core loci.");

f. Yan Ma, et. al., *Next generation sequencing: Improved resolution for paternal/ maternal duos analysis,* FORENSIC SCIENCE INTERNATIONAL: GENETICS 24 83–85 (2016) ("The Illumina MiSeq FGx Forensic Genomics System became available in early 2015 and has been validated by the manufacturer. It will save a significant amount of time and number of assays needed to achieve the same genetic information. Furthermore, NGS-based methods can obtain not only the length based genotypes but also the full sequence data from all STR loci. The availability of full sequence information from NGS makes it possible to investigate the true variation (SNPs and InDels) within STR loci and identify previously unknown alleles and mutational events in kinship analysis.");

g. Kristiaan J. van der Gaaga, et. al, *Massively parallel sequencing of short tandem repeats—Population data and mixture analysis results for the PowerSeqTM system*, FORENSIC SCIENCE INTERNATIONAL: GENETICS 24 86–96 (2016) ("As was expected,

MPS STR genotyping revealed substantial genetic variation in addition to the variation in repeat length that is detected using CE.");

h. Katherine Butler Gettings, *Sequence variation of 22 autosomal STR loci detected by next generation sequencing*, FORENSIC SCIENCE INTERNATIONAL: GENETICS 21 15–21 (2016) (NIST study) ("Industry competition has led to drastic drops in sequencing costs in recent years, and advances in library preparation methods and read lengths now enable sequencing of forensic STR loci, as demonstrated by several laboratories. . . . The sequences obtained in this limited data set give an indication of the level of diversity expected in the larger population and provide examples of how isoalleles (alleles of the same length but different sequence) can improve discrimination and mixture deconvolution in forensic casework.");

i. Nicole M.M. Novroskia, *Characterization of genetic sequence variation of 58 STR loci in four major population groups*, FORENSIC SCIENCE INTERNATIONAL: GENETICS 25 214–226 (2016) ("Overall, the sequence variation observed herein and those reported in the literature show similar trends and indicate the diversity of sequence variation yet to be uncovered in larger datasets.");

j. Walther Parson, et. al., *Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements*, FORENSIC SCIENCE INTERNATIONAL: GENETICS 22 54–63 (2016) ("Massively parallel sequencing (MPS) is adding a new dimension to the field of forensic genetics, providing distinct advantages over CE systems in terms of captured information, multiplex sizes, and analyzing highly degraded

samples. In recent years, MPS has been applied to the generation of STR sequence data with the general outcome that STRs can be successfully typed producing genotypes compatible with those of CE analyses, even from compromised forensic samples. Furthermore, MPS derived STR genotypes provide additional information to that generated by CE separation by capturing the full nucleotide sequence underlying the repeat units and nearby flanking regions.");

k. Nouran Almalki , et. al., *Systematic assessment of the performance of Illumina's MiSeq FGxTM forensic genomics system*, ELECTROPHORESIS 38 846–854 2017 ("Currently, DNA length polymorphisms of short tandem repeats (STRs) are used for forensic identification of individuals. This technique, which utilizes the polymerase chain reaction (PCR) and capillary electrophoresis (CE), has two major limitations. First, it does not sequence STR amplicons. Consequently, it fails to identify sequence variations within amplicons of equal size that could be used to improve the STR system's discriminatory power. Second, the number of loci used for individual identification, which will reach 35 STRs (24 aSTRs & 11 Y-STRs) with the Spectrum CE system that will be introduced soon, is nearing its limit due to the technical constraints of CE. In contrast, massively parallel sequencing (MPS) offers the opportunity to sequence each locus and simultaneously increase the number of STRs evaluated, particularly by including more loci on the Y- and X-chromosomes.");

l. Press Release, *FBI Approves Verogen's Next-Gen Forensic DNA Technology for National DNA Index System*, ASSOCIATED PRESS (May 2, 2019), https://apnews.com/press-release/pr-businesswire/e71b0cc65a4d4a14817dd07046d6ea3a

("Verogen Inc., the world's first DNA sequencing company dedicated to forensic science, announced today that the United States Federal Bureau of Investigation has approved the company's next-gen forensic DNA technology for use by forensic laboratories generating DNA profiles for upload to the National DNA Index System (NDIS).").

As these articles suggest, there are now "[t]urnkey implementation services to get [forensic laboratories] up and running with massively parallel sequencing, from start to finish." BATTELLE CORPORATION, *Forensic Genomics*.[29]

In light of these services and the other developments discussed above, the FBI will be hard pressed to claim that sequence variants are relatively uncommon, or that sequencing is too time consuming or costly, or that it has practical limitations, or that it is rarely done. On the contrary, the foregoing scientific developments illustrate that the forensic DNA community adheres to the basic scientific principle that DNA matching techniques must not be based on superficially attractive, but ultimately incomplete and therefore unscientific comparisons that ignore readily available information.

The principle itself is not unique, having been articulated well in an article published long ago:

> Unfortunately, when [ABO and VNTR alleles having the same length but differant sequence] are observed in forensic analyses, they are generally treated as equivalents of their more frequent equal-sized bands. If this occurs, a declaration of a 'band match' may erroneously be considered as proof of a 'genetic match.'
>
> More recently, the genetic traits analyzed in forensic situations have included the type of noncoding DNA labeled 'short tandem repeats' (STRs). One such

[29] https://www.battelle.org/government-offerings/homeland-security-public-safety/security-law-enforcement/forensic-genomics (last visited Jan. 3, 2021).

> example is HUMTHO1, which typically has from 6 to 14 repeats of a core sequence consisting of AATG. Here too, variations in nucleotide sequence have been found among like-sized repeats. . . . Thus as was to be expected, STRs present the same problem as [ABO and] VNTRs when analyzing their repeat number versus their repeat sequences.
>
> In addition to the preceding examples, there are numerous other situations in which laboratory test results seeking to match suspect with evidence have been questioned. The need to be as certain as possible in genetically characterizing a particular individual becomes ever more necessary as genetic traits are increasingly used as circumstantial evidence in criminal cases. Juries, and even judges, can become confused if, in addition to the complexities of the scientific procedures, challenges can be raised as to the interpretation of the test results.
>
> Can the need to identify avoid the pitfalls produced by our ever increasing ability to detect genetic variability? We believe that the most promising solution lies in DNA base sequence analysis of both evidence and suspect. The technique involved is well established.

L. Levine & L. Kobilinsky, *The need to identify: a word of caution*, 101 GENETICA 141–144 (1997).

As these authors stress, in a situation where the length and sequence of a genetic chracteristic may vary, "caution must be observed in attempting to link a suspect to the evidence. Additional laboratory tests must be conducted in order to ascertain the genotypes involved. It must be emphasized that a genetic match for any one trait does not, by any means, permit one to declare that a suspect is the source of a particular piece of biological evidence. . . . In addition, it should be clearly understood that the constellation of matched genotypes, forming what is commonly called a 'genetic profile', is not individual-specific. The probability that some other person has the same genetic profile is determined by calculating the product of the frequencies of the individual genotypes in the relevant geographic population. However, such a calculation can only be

done if all examined genotypes have been shown to be matched. The lack of a genetic match for even a single trait excludes the individual as a suspect." *Id.* at 142.

The authors' reminder that the length versus sequence problem goes back at least to the days of ABO blood typing has an important link to modern legal developments. Recent statistics on exonerations highlight the danger of uncritically accepting claimed consistencies in forensic comparison evidence that are not grounded in the most up-to-date developments in science. For example, it was reported in 2011 as follows:

> A study of 200 DNA exonerations found that expert testimony (55%) was the second leading type of evidence (after eyewitness identifications, 79%) used in the wrongful conviction cases. Pre-DNA serology of blood and semen evidence was the most commonly used technique (79 cases). Next came hair evidence (43 cases), soil comparison (5 cases), DNA tests (3 cases), bite mark evidence (3 cases), fingerprint evidence (2 cases), dog scent (2 cases), spectrographic voice evidence (1 case), shoe prints (1 case), and fibers (1 case). Brandon L. Garrett, Judging Innocence, 108 Colum. L. Rev. 55, 81 (2008). These data do not necessarily mean that the forensic evidence was improperly used. For example, serological testing at the time of many of these convictions was simply not as discriminating as DNA profiling. Consequently, a person could be included using these serological tests but be excluded by DNA analysis. Yet, some evidence was clearly misused.

Paul Giannelli, Edward Imwinkelried and Joseph Peterson, *Reference Guide on Forensic Identification Expertise*, FEDERAL JUDICIAL CENTER, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE: THIRD EDITION 55, 62 n.32 (2011). The lesson learned from these exonerations by three of the nation's leading evidence scholars was that "[t]hese data do not necessarily mean that the forensic evidence was improperly used. *For example, serological testing at the time of many of these convictions was simply not as discriminating as DNA profiling. Consequently, a person could be included using these*

serological tests but be excluded by DNA analysis. Yet, some evidence was clearly misused." (Emphasis added.)

It must be remembered that for years serological comparison evidence was considered the gold standard of forensic comparative evidence. Courts freely admitted this evidence and juries routinely relied on it in convicting defendants and in some cases in imposing the death penalty. Now, thanks to first generation DNA typing techniques, we are learning that in a not insignificant number of cases the consistencies that had been testified to by the confident practioners of this forensic discipline were in fact false.

Given the recent development of NGS capability, the same history will no doubt repeat itself with STR testing. There is no dispute that if an evidentiary crime scene sample and the defendant's sample are identical in length at thirteen loci, but differ by sequence at one loci, the defendant is excluded as the source of the evidentiary sample. It will not be too long, therefore, before we will start seeing exonerations based on sequence DNA testing in which the STR evidence at trial showed a consistency. The explanation for such results by the evidence scholars will be exactly the same: "[t]hese data do not necessarily mean that the forensic evidence was improperly used. For example, [STR] testing at the time of many of these convictions was simply not as discriminating as DNA [sequence] profiling. Consequently, a person could be included using these [STR] tests but be excluded by [NGS] analysis."

The question, then, is whether it is fair and just under the due process clause, reliable under *Daubert* and generally accepted by the scientific and legal community to continue to use STR testing when all scientists now agree (1) STR testing is simply not as

discrminating as DNA sequence testing; and (2) we have the capability and resources right now to conduct sequence testing of forensic samples. Mr. Bowers suggests that the answer to that question should be guided by a general principle of law, which holds that "a consistency produced by ignoring individual differences is a false consistency." *Eddings v. Oklahoma*, 455 U.S. 104, 113 (1982).

5) Interpretation Issues With Complex DNA Mixtures

Forensic DNA samples from crime scenes often contain DNA from more than one individual. Such a DNA profile representing two or more contributors is termed a DNA mixture. Butler, Fundamentals at 320. An analyst knows that they are dealing with a DNA mixture, versus a single source sample, if they observe more than two alleles at two or more loci, or if loci with only two alleles display significant peak height imbalance.[30] Butler, Advanced Topics at 129. Unlike the straightforward analysis involved in interpreting a high quality single source DNA profile, mixtures are often ambiguous, and the process of interpreting them can be highly subjective. In particular, mixtures which cannot be resolved into single source components ("indistinguishable" mixtures)[31]

[30] Two alleles from the same contributor should be roughly the same height, within a degree of tolerance (called a "peak height ratio" (PHR)). If the height of two allelic peaks observed at a given locus are not within this predetermined tolerance—*i.e.*, they are "imbalanced"—this is a sign that they actually originate from two people rather than one.

[31] Mixtures may sometimes be resolved or 'deduced' into individual sources based on the relative amounts of DNA contributed by each source; the source contributing more DNA is the 'major contributor,' and the source(s) contributing less DNA is the 'minor contributor.' Analysts use the height of allelic peaks on the electropherogram as a proxy for how much DNA is originating from each contributor. Laboratories have specific criteria regarding how much difference they have to observe between peak heights to pull out a major profile; it would be inappropriate to 'eyeball' a mixture to determine whether it impressionistically appears that there is 'enough' of a difference between contributors to deduce a major profile. Some mixtures

involve a great deal of subjective decision making. Studies have shown that subjective interpretation of indistinguishable DNA mixtures can lead to widely divergent results from one analyst to the next, even analysts in the same laboratory applying the same set of protocols. *See* Dror and Hampikian, *Subjectivity and bias in forensic DNA mixture interpretation*, SCI. & JUSTICE 51(4) 204–208 (2011);[32] NIST Interlaboratory Mixture Interpretation Study 2013 ("MIX13") (discussed *infra*) and forerunner NIST mixture studies (e.g. MIX05). John M. Butler , Margaret C. Kline , Michael D. Coble, *NIST interlaboratory studies involving DNA mixtures (MIX05 and MIX13): Variation observed and lessons learned*, FORENSIC SCIENCE INTERNATIONAL: GENETICS 37 (2018) 81-94, available at https://www.fsigenetics.com/action/showPdf?pii=S1872-4973%2818%2930248-5 (last visited Jan. 4, 2021). Two complicating factors in mixture interpretation are: (a) "the potential for allele stacking", and (b) "potential alleles in the stutter position." Advanced Topics at 153.

a. The Problem Of Allele Stacking

As described *supra*, when an individual has two of the same alleles at a locus (*i.e.* is a homozygote), that person's alleles "stack" on top of one another and present as a

encountered in casework do not meet these criteria and therefore the mixture must be treated in its totality rather than as individual single source profiles.

[32] In this study, seventeen examiners from one government laboratory were provided a mixed DNA profile from a sex assault case and asked to interpret the profile and compare it to a suspect's reference profile. The original case work analyst had determined that the suspect could not be excluded as a contributor to the mixture. The seventeen examiners came to a variety of conclusions: 1 concluded "cannot exclude"; 12 "excluded" and 4 deemed the results "inconclusive." Among other things, these results underscore the subjectivity of complex mixture interpretation. *Available at* http://www.scienceandjusticejournal.com/article/S1355-0306(11)00096-7/pdf (last visited Jan. 3, 2021).

single peak on the electropherogram. Similarly, when multiple contributors to a DNA mixture possess the same allele at a locus, those alleles also "stack" and present as a single peak. *See, e.g.*, Figure 2, below. This is known as allele stacking or allele sharing. There are two important consequences of allele stacking. One consequence is that "allele sharing makes accurately deducing the number of contributors to a mixture challenging – and the challenge only grows with each additional contributor to a DNA mixture." Butler, Advanced Topics at 169. If an analyst cannot accurately determine how many contributors there may be in a mixture, the analyst cannot reliably interpret the mixture. Studies have shown that, because of allelic stacking, more than 75% of known four-person mixtures would be misclassified as two- or three-person mixtures based on the maximum number of alleles detected at any given locus. Paoletti et al., *Empirical Analysis of the STR Profiles Resulting from Conceptual Mixtures*, J FORENSIC SCI 50(6), 1361–66 (Nov. 2005).[33] Inaccurate interpretation of the mixture impacts whether an individual is included or excluded as a potential contributor to the mixture, as well as the associated statistical analysis. Butler, Advanced Topics at 335 ("[S]ome of these genotype combinations may not fit a reasonable interpretation of the data" depending on the actual number of contributors present.).

---

33 https://projects.nfstc.org/workshops/resources/literature/Genetic_Analysis_and_Interpretation/14_Empirical%20Analysis%20of%20the%20STR%20Profiles%20Resulting.pdf (last visited Jan. 3, 2021).

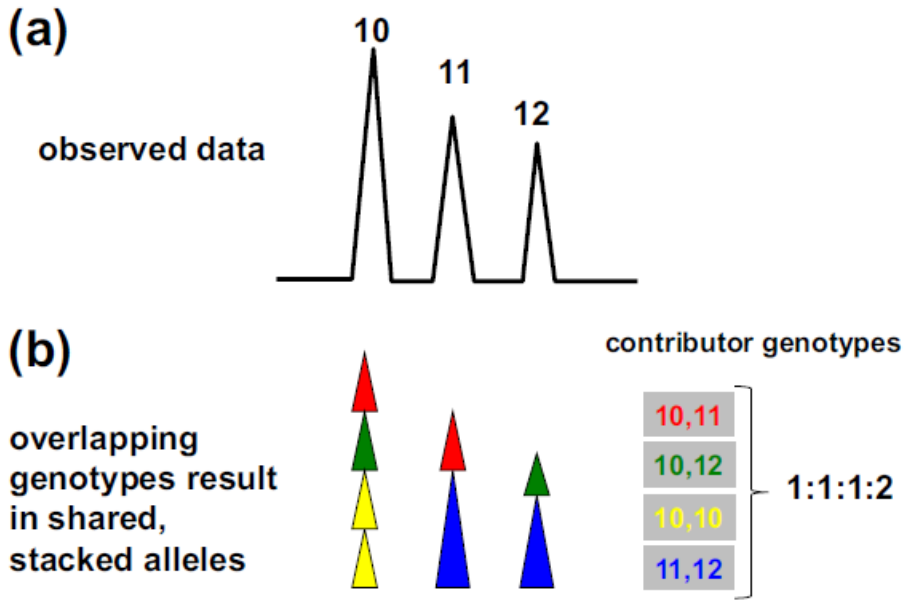

Figure 2. Hypothetical mixture which (a) exhibits only three alleles at a locus (and is thus suggestive of a two person mixture), (b) is actually comprised of three low level contributors plus a single higher level contributor, whose alleles stack on top of one another. Figure from *Interpretation*, at 160, Fig. 7.1.

A second consequence of allele stacking is that it diminishes the utility of the stochastic threshold as a means of determining whether allelic dropout has occurred. Butler, Advanced Topics at 163 ("[T]he potential of allelic stacking, especially with more than two contributors, can limit the usefulness of a stochastic threshold."). When the stochastic threshold cannot be effectively utilized, interpretation and statistical analysis are not reliable. Reliance on a stochastic threshold without considering the possibility of allelic dropout may result in a false inclusion or exclusion.

A stochastic threshold is a Y-axis value on the electropherogram (measured in relative fluorescence units, or RFUs). The value is established by the laboratory's internal

validation studies.[34] Data below the stochastic threshold is in the "potential 'danger zone' of unreliable results." John M. Butler and Carolyn R. Hill, *Scientific Issues with Analysis of Low Amounts of DNA*, NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, BIOCHEMICAL SCIENCE DIVISION (2010).[35] When peaks from the evidence sample fall below the stochastic threshold at a locus, there is a risk that allelic dropout—or loss of genetic data—is occurring at that locus. Specifically, allelic dropout occurs when only one of a DNA contributor's two alleles at a given locus is detected by the DNA typing process.[36] This is a common problem associated with low template DNA analysis. *See United States v. McCluskey,* 954 F. Supp. 2d 1224 (D.N.M. 2013) (results of low copy number (LCN) DNA testing were not sufficiently reliable to be admissible under *Daubert*).

As described *supra,* allelic dropout can happen when an individual's DNA is present in low levels, is degraded, or is inhibited. Stochastic thresholds are important because seeing an allele below the stochastic threshold at a given locus is a "warning indicator" that the partner allele (*i.e.*, the second allele of the pair) may have dropped out. *Id.* at 163–64. Dropout of a partner allele could lead an analyst to detect a "false homozygote." *Id*. For example, if the true contributor of an evidentiary DNA sample possesses an 8 and a 12 allele at a given locus, but due to allelic dropout only the 8 allele

---

[34] The stochastic threshold value(s) should be a part of the lab's written protocols as determined by the lab's validation.

[35] https://www.promega.com/resources/profiles-in-dna/2010/scientific-issues-with-analysis-of-low-amounts-of-dna/ (last visited Jan. 3, 2021).

[36] There can also be loss of both alleles at a given location, which is called locus dropout.

is detected, an individual who is homozygous for the 8 allele (*i.e.*, possesses two 8 alleles) could be falsely implicated, while the true contributor could be falsely excluded. *See, e.g.*, Sec. II.D, Fig. 3 (discussion of Case 5 in MIX13 study). While the stochastic threshold serves some purpose, in that DNA data that is unambiguously in the stochastic range (*i.e.*, below the stochastic threshold) is clearly at risk of being incomplete, DNA data that rises above the stochastic threshold is not necessarily safe. This is especially true with complex DNA mixtures, due to the potential for contributors to share alleles (*i.e.*, allele stacking).

Allele stacking makes over-reliance on using the stochastic threshold to determine whether or not drop-out has occurred particularly dangerous with mixtures. "Just because allelic peaks at a locus are above an established stochastic threshold does not mean that no allele drop-out has occurred in a complex mixture." Butler, Advanced Topics at 163–64. Allele stacking can falsely elevate a peak at a particular locus above the stochastic threshold. When two or more sub-threshold alleles stack on top of each other, they may present as a peak that surpasses the stochastic threshold, which in turn may give the false impression that the DNA at that locus is free from the risk of allelic dropout and can be confidently interpreted. In reality, however, each contributor to the falsely elevated peak is in the stochastic (dropout) zone. *See, e.g.*, Figure 2 (showing peaks from multiple low level contributors stacking upon one another and presenting as three relatively tall peaks). "The concept of a stochastic threshold can become meaningless in complex mixtures due to the potential for allele stacking." Butler, Advanced Topics at 94–95; *id.* at 177

("stochastic thresholds often lose their value and meaning when allele sharing is possible with three or more contributors to a DNA mixture").

While some level of allelic stacking will occur in any DNA mixture, there is no objective way to determine whether allelic stacking is occurring at any given locus in an indistinguishable DNA mixture profile because there is no way to tell whether an observed peak comes from one contributor, or actually is the combined low level (*i.e.*, sub-stochastic) contributions of two or more individuals. *See, e.g.*, *supra,* Figure 2 (the 10 allele demonstrates how an above-threshold peak can originate from a combination of two or more individuals whose individual contributions are below the stochastic threshold). Analysts typically make educated guesses based on other information present in the DNA profile because there are no objective guidelines or protocols to guide the analyst's determination that allele stacking is or is not occurring. These educated guesses may or may not lead to accurate conclusions. The more complex the mixture the more difficult it is to make a reliable educated guess.[37] Ultimately, and unavoidably, "allele drop-out and potential allele sharing from multiple contributors lead to greater uncertainty in the specific genotype combinations that can be reliably assumed." Butler, Advanced Topics at 177. And, as Dr. Butler has unambiguously warned, "[w]hen there is a high degree of interpretation uncertainty from an evidentiary sample, it makes little sense to try and draw conclusions . . . and expect those conclusions to be reliable." *Id.*

[37] This is why most labs have a "complexity rule" that precludes interpretation of samples that are too complex either because of the number of contributors or the potential for allelic drop-out.

Nevertheless, the analyst, in this case, chose to interpret these complex indistinguishable mixtures.[38]

b. The Heightened Challenge Of Distinguishing Artifacts From Real DNA In Complex Mixtures

Another significant source of uncertainty in mixture interpretation is distinguishing real alleles from artifacts, particularly in an extremely common artifact known as "stutter." Stutter is a by-product of the amplification (*i.e.*, copying) step in the DNA testing process, and typically results in a small peak one repeat less than its parent "true allelic" peak (*e.g.*, the process would produce a smaller "stutter" peak in the 7 allele position when there is a true 8 allele at that locus). "Because stutter products are the same length as actual allele PCR products, it can be challenging to determine whether a small peak is a real allele from a minor contributor[39] of the original sample or a stutter product of an adjacent allele created during the PCR amplification process." Butler, Advanced Topics at 76. When there are one or more minor contributors present whose alleles are similar in height to the stutter peaks, this task is not just "challenging," it is impossible.

[38] Allelic dropout is not simply a theoretical possibility. It is "ever-present" and a "real issue faced with complex mixtures" because "[s]ensitive DNA detection technology has the potential to outpace reliable interpretation." Butler, Advanced Topics at 174, 177. "If a laboratory desires to develop appropriate protocols that will enable reliable interpretation of DNA from low-level DNA or mixtures involving three or more contributors, then validation studies need to be performed with known samples that mimic the amounts of DNA and complexity of profiles where stochastic effects and allele dropout are expected." *Id.* at 164. The 2017 SWGDAM Interpretation Guidelines require internal validation studies to establish the stochastic threshold, while acknowledging that reliance on the stochastic threshold may not be appropriate in mixture samples where allele sharing is possible. SWGDAM Interpretation Guidelines (2017) 1.7, 1.7.1, and 1.7.1.3.

[39] A minor contributor is simply an individual contributing a smaller amount of DNA (which will appear on the electropherogram as smaller peaks) relative to other contributors to a DNA mixture.

Id. at 58–59. When there is an optimal amount of DNA present, stutter peaks tend not to exceed a certain height relative to the associated parent "true allelic" peak. However, the fact that a low-level peak is adjacent to a larger peak does not necessarily mean that it is stutter. *Id.* at 142 ("It is not always possible to exclude stutter since they are allelic products and differ from their associated allele by a single repeat unit").

For complex mixtures, stutter is even more problematic. Not only does it become impossible to distinguish real DNA from stutter, but stutter peaks can stack in exactly the same way real allelic peaks do. *Id.* at 71. Thus, stutter can stack on a sub-threshold allelic peak and present as a peak that artificially surpasses the stochastic threshold. Moreover, with a mixture containing one or more low level contributors, "higher levels of stochastic variation can lead to more variability in peak height ratios of heterozygotes and more significant stutter products." *Id.* at 160. In other words, when there are low-level DNA contributors present in a mixture, stutter peak heights can exceed expected values (*i.e.*, the values set by validation studies) and be confused with real allelic peaks. In fact, in low template DNA samples stutter peaks may often exceed their "parent" peak making distinguishing a true peak from an artifact impossible. Therefore, with complex mixtures, "[t]his variation leads to a lower confidence in appropriately allocating allele pairs into individual contributor genotypes with complex mixtures." *Id.*

c. Uninterpretable Mixtures And Probabilistic Genotyping

As explained above, low level DNA samples and DNA mixtures of two or more contributors pose a problem to DNA forensic analysts. In the past analysts dealt with this challenge by calculating statistics concerning the probability of inclusion. But these

statistics were general in nature and continue to be the subject of much controversy. *See* William C. Thompson, Laurence D. Mueller, and Dan E. Krane, *Forensic DNA Statistics: Still Controversial in Some Cases*, THE CHAMPION, December 2012, at 12–23. Recently, labs have begun using software programs to analyze complex DNA mixtures. Probabilistic genotyping software programs, such as STRmix, are designed to calculate a statistic to contributors of such mixtures when one could not be determined in the past. These programs use biological modeling, statistical theory, computer algorithms, and probability distributions to calculate likelihood ratios (LRs). LRs are the statistic calculated by these probabilistic programs, which reflects the relative probability of a particular finding under alternative theories about its origin. *Id*. In forensic DNA analysis, that LR can be stated as the profile is X amount of times more likely if the defendant and a certain number of other unknown, unrelated contributors contributed to the mixture.

1. What Is STRmix?

STRmix, the probabilistic genotyping software used in this case, was developed by the Institute for Environmental Science and Research (ESR) in New Zealand. It uses computer science algorithms to perform "complex" mathematical and statistical calculations. *See* Jo-Anne Bright, Duncan Taylor, et al., *Developmental validation of STRmix, expert software for the interpretation of forensic DNA profiles*, FORENSIC SCIENCE INT'L: GENETICS 23:226–239, 227 (July 2016) ("Developmental validation of STRmix"). Indeed, STRmix is a software program—it is not used for any of the other steps of DNA analysis described above. *See generally id.* It is only after a DNA analyst in

a laboratory performs the regular steps developing a DNA profile from a sample that STRmix adds an additional step. .It is performed not in a lab by a trained scientist but instead by a person sitting at a computer screen who runs a complex computer software program, which he or she may not completely understand. This program is designed to answer the classic question in forensic DNA interpretation: what are the profiles of the contributors to this mixture?

The program relies on analysts to collect the data by reviewing the electopherograms ("epg") developed in a case and discarding the peaks below the lab's analytic threshold. *See* Duncan Taylor, Jo-Anne Bright, and John Buckleton, *The Interpretation of Single Source and Mixed DNA Profiles*, FORENSIC SCI INTL: GENETICS 7 519 (2013). Artifacts like pull-up and forward stutter are also removed manually. *Id*.

The backbone of the STRmix software system is a computing algorithm called the Markov Chain Monte Carlo ("MCMC") method of calculating probable outcomes. *Id*. at 233. The implementation of the MCMC algorithm in STRmix utilizes statistical models to simulate hypothetical true alleles while incorporating stochastic effects. *Id*. It then assesses those simulated alleles and makes conclusions about what is true DNA as opposed to artifacts in a sample. *Id.* Based on those conclusions, the likelihood ratio is then generated as a further statistical assumption. The reason that MCMC is used is that there are an exponentially enormous number of combinations of assumptions and outcomes that arise from any mixed sample. It would be practically impossible to do such a calculation without a computer running sophisticated software.

2. How Does STRmix Compare To Other Probabilistic Software Programs?

There is no agreement within the forensic community about which probabilistic software programs or methods to employ, if any, when analyzing low template DNA or complex mixture samples. There are at least eight different probabilistic genotyping software programs in the country. A brief discussion of two others, the Forensic Statistical Tool (FST) and TrueAllele, illustrates their differences from STRMix. These two programs vary from STRmix and each other in the manner in which they collect data, the necessary assumptions they make to perform their statistical calculations, and the actual underlying mathematical principles used to make these calculations.

The Forensic Statistical Tool (FST) was developed in-house at the Office of Chief Medical Examiner (OCME) of the City of New York by Dr. Theresa Caragine and Dr. Adele Mitchell and is used in all complex mixture cases in New York City. Similar to STRmix, FST relies on analysts to collect the data used in the calculations and analysis. An analyst reviews the electropherogram and determines whether alleles are present at each locus by utilizing the lab's analytic threshold. The analyst inputs this information, along with a known suspect profile, into the FST software. The analyst then sets the parameters for running the program including whether the mixture contains two or three contributors. The software then outputs a "Forensic Statistic Comparison Report," summarizing the data that was input and indicating the resultant likelihood ratio. FST differs from STRmix in how it calculates the LR. Unlike STRmix, FST does not use MCMC algorithms in making these calculations, instead relying only on Bayesian

statistics. Bayesian statistics describe the probability of an event, based on conditions that might be related to the event. *See* John Butler, FORENSIC DNA TYPING: BIOLOGY, TECHNOLOGY, AND GENETICS OF STR MARKERS 459 (2d ed. 2005). FST also calculates the "drop-out" rate differently than STRmix. FST calculates the allelic "drop-out" based on the quantitation values of given DNA samples, rather than peak height variation, as STRmix does.

TrueAllele differs significantly from STRmix and FST in the manner in which it collects, interprets, and calculates the data. TrueAllele has been developed by Cybergenetics of Pittsburgh, Pennsylvania under the direction of Dr. Mark Perlin. TrueAllele is a fully continuous probabilistic approach that analyzes the epgs and considers the genotypes at every locus of each contributor, taking into consideration the mixture weights of the contributors, the DNA template mass, polymerase chain reaction (PCR) stutter, relative amplification, DNA degradation, and the uncertainties of all these variables.

Unlike FST and STRmix, TrueAllele does not rely on an analyst's interpretation of what constitutes a true allele by using analytical thresholds dictated by laboratory protocol in order to collect its data. TrueAllele instead considers all the data present in the sample, even those peaks below the lab's analytic threshold. In essence, the calculations made by TrueAllele are based upon more information than used by FST and STRmix. Unlike FST, TrueAllele accounts for "drop-out" rates as a function of peak heights and peak height ratios seen in the sample rather than based on the quantity of DNA in the sample. Like STRmix, but unlike FST, it uses MCMC algorithms to calculate likelihood

function that compares genotypes relative to a population and computes a match LR. *See People v. Wakefield*, 47 Misc.3d 850, 859 (N.Y. Sup. Ct. 2015).

6) DNA Mixture Evidence Challenged in this Case

Although this Motion challenges the admission of all of the government's DNA evidence on various grounds, the DNA Report alleges that [redacted] These are the items with alleged DNA mixtures and STRmix results that the defense seeks to exclude through this Motion on the ground that the STRmix genotyping software is unreliable.

Agent Plaza's DNA report summarizes her conclusions about [redacted]. For example, [redacted]

[redacted]

An alternative hypothesis is also posited that [redacted]

[redacted]

IV. ARGUMENT

A. This Court should order the immediate discosure of the FBI's DNA testing litigation packets in order to allow Mr. Bowers to fully and fairly litigate this *Daubert* motion.

The forensic DNA community has a long history of requiring DNA testing laboratories to open its work product to the scrutiny of the scientific community. For instance, the first NRC Report on DNA Technology in 1992 stated emphatically that "[a]ll data and laboratory records generated by analysis of DNA samples should be made freely available to all parties" because "[s]uch access is essential for evaluating the analysis." NRC I at 150. The NRC I explained:

> The court and the jury should have no reason to doubt the accuracy of the processing of information. Laboratories and experts have a particular responsibility to ensure that they are open and candid with the courts. Any reservations about inadequacies or errors should be promptly revealed, and failure to do that should be dealt with seriously. The court should not hesitate to exercise contempt powers and exclude experts who have misled deliberately in the past. Private trade associations and other appropriate groups should also apply pressure to ensure accuracy and candor.
>
> Protective orders should not be used to prevent experts on either side from obtaining all relevant information, which can include original materials, data sheets, software protocols, and information about unpublished databanks. A protective order might be appropriate to limit disclosures by attorneys and experts to third parties about proprietary information acquired in the course of a particular case; but as a general rule, any scientific information used in a case should be open to widespread scientific scrutiny.

NRC I at 148.

The NRC II agreed with this recommendation:

> The 1992 National Research Council (NRC) report stated that "all data and laboratory records generated by analysis of DNA samples should be made

> freely available to all parties," and it explained that "all relevant information . . . can include original materials, data sheets, software protocols, and information about unpublished databanks" (NRC 1992, p 150, 148). Certainly, there are no strictly scientific justifications for withholding information in the discovery process, and in Chapter 3 we discussed the importance of full, written documentation of all aspects of DNA laboratory operations. Such documentation would facilitate technical review of laboratory work, both within the laboratory and by outside experts....Our recommendation that all aspects of DNA testing be fully documented is most valuable when this documentation is discoverable in advance of trial.

NRC II at 168–69.

The President's Council of Advisors on Science and Technology (PCAST) agrees: "Because errors due to human failures will dominate the chance of coincidental matches, the scientific criteria for validity as applied require that an expert . . . (2) should routinely disclose in reports and testimony whether, when performing the examination, he or she was aware of any facts of the case that might influence the conclusion, and (3) should disclose, upon request, all information about quality testing and quality issues in his or her laboratory." President's Council of Advisors on Science and Technology, REPORT TO THE PRESIDENT: FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS 75 (Sept. 2016), available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf (last visited Jan. 3, 2021).

Courts have specifically held that the type of electronic and other litigation packet data requested by the defense in this case is discoverable. *See, e.g.*, *United States v. DeLeon*, 426 F. Supp. 3d 1034, 1067 (D.N.M. 2019) ("In meeting obligations under Rule 16(a)(1)(E), (F), and (G), the Jencks Act, and *Brady/Giglio*, and to comply with the

Department's policies of broad disclosure, the prosecutor should be attuned to the following four steps: . . . Third, if requested by the defense, the prosecutor should provide the defense with a copy of, or access to, the laboratory or forensic expert's 'case file,' either in electronic or hard-copy form. This information, which may be kept in an actual file or may be compiled by the forensic expert, normally will describe the facts or data considered by the forensic expert, include the underlying documentation of the examination or analysis performed, and contain the material necessary for another examiner to understand the expert's report. The exact material contained in a case file varies depending on the type of forensic analysis performed. It may include such items as a chain-of-custody log; photographs of physical evidence; analysts' worksheets or bench notes; a scope of work; an examination plan; and data, charts and graphs that illustrate the results of the tests conducted."); *United States v. Willock*, 696 F. Supp. 2d 536, 578 (D. Md. 2010) ("The failure to produce the documents constituting the bases and reasons for [the expert's opinions] was particularly troublesome in this case because . . . there is a substantial debate within the scientific community, as well as the Courts, regarding the degree to which firearms toolmark identification evidence passes muster under Fed.R.Evid. 702 and Daubert."); *United States v Yee*, 129 F.R.D. 629 (N.D. Ohio 1990) (even before 1993 amendment to Federal Rule of Criminal Procedure 16(a)(1)(E), a federal magistrate judge granted discovery of matching criteria, environmental insult studies, population data, and proficiency tests as "predicate materials" essential to the defense in a DNA-testing case); *People v. Seepersad* 58 Misc. 3d 1227 (N.Y. Sup. Ct. 2018) ("[T]he prosecution and its agents are constitutionally obliged to disclose arguably

exculpatory information to the defense. This court does not believe that [electronic] data critical to the assessment of a laboratory's inculpatory [DNA] conclusions could be outside the rule, particularly given that different software used to analyze the data will yield different results."); *People v. Gills,* 52 Misc. 3d 903, 908–09 (N.Y. Sup. Ct. 2016) ("Because the electronic raw data is subject to interpretation by both the software program which processes it and the analyst who constructs the DNA profile, the defendant cannot effectively challenge the DNA evidence without it."); *People v Davis* 601 N.Y.S.2d 174 (N.Y. App. Div. 1993) (Lifecodes was required on constitutional grounds to turn over statistical data underlying a DNA probability estimate); *State v Schwartz*, 447 N.W.2d 422, 427 (Minn. 1989) ("[A]ccess to the data, methodology, and actual results is crucial so a defendant has at least an opportunity for independent expert review."); *id.* (although a laboratory disclosed its protocol, laboratory notes, autoradiographs, and frequency tables, its refusal to supply "more specific information on its methodology and population data base" was a reason to exclude the findings). *See* ABA Criminal Justice Standards on DNA Evidence, Standard 4.1(a)(viii) (stating the prosecution should disclose "all raw electronic data produced during testing").

Because the government has refused in a timely manner to share data that is necessary for Mr. Bowers' experts to review in order to assess the reliability of the FBI's testing results, the appropriate remedy is to order immediate discovery or to disallow the government's DNA testing results on the ground that its refusal to disclose this data has prevented Mr. Bowers from attacking the reliability of this evidence. *See United States v. Cerna*, No. CR 08-0730 WHA, 2010 WL 11627594, at *9 (N.D. Cal. Dec. 17, 2010)

("Secret bases for expert testimony are inherently unreliable. There is no way to verify that they actually exist, much less that they are reliable."); *People v. Fortin*, 218 Cal. Rptr. 3d 867, 874 (Cal. App. Dep't Super. Ct. 2017) (trial court properly refused to admit the results of the Abel "Sexual Interest Assessment" test under *Daubert* analysis, in part because "[c]ross-examination would be thwarted by Dr. Flores's inability to explain how Fortin responded to the photo display and what this signifies. The process of analyzing responses is closely-guarded proprietary information that Dr. Abel refuses to share.").

The defense is particularly concerned that the government's refusal to disclose the DNA litigation packets, which would include the specific GlobalFiler profiles for Mr. Bowers and for any comparison samples, is preventing the defense from fully determining whether this case presents the problems of null allelles, sequencing versus length variation, and complex mixture interpretation discussed above.

In order to fully present why the FBI's conclusions are unreliable and unscientific, the defense seeks the undelying testing data before a *Daubert* hearing because it is the government that will be required to demonstrate that its length-based testing is reliable and why and how STRmix™ meets the standards for admissability. STRmix™ is a software program that is in its relative infancy of use in criminal cases in the United States. Moreover, it has been excluded under the *Daubert* standard before. Recently, in *United States v. Gissantaner*, 417 F. Supp. 3d 857 (W. D. Mich. 2019), a District Court Judge in the Western District of Michigan, after holding extensive *Daubert* hearings over a year and half long period, involving several expert witnesses including two Court-appointed experts who provided independent written reports to the Court, excluded

STRmix™ results due to insufficient internal validations of its use in casework by the Michigan State Patrol Laboratory.

The defense believes that similar lengthy litigation will be required to address the complex challenges it expects to mount based on the weakness of the DNA evidence in this case. The government will carry the burden to prove that its length-based STR testing and its STRmix™ LR statistics are relevant and reliable; and more probative than prejudicial to be used as evidence in this case. The government will need to prove that their own internal validations for the use of STRmix™ in the FBI Laboratory case work do not result in an intolerably high degree of error. Given the weakness of many of the conclusory results disclosed to date, the defense anticipates that the government will fail to carry this burden because its statistics are not based on a reliable methodology and its reliance on "Verbal Equivalencies" to support these statistics do not meet the scientific or technical requirements outlined in *Daubert* and FRE 702. Finally, the use of weak STRmix™-generated LR statistics in this case are not generally accepted in the broader scientific community, and defense does not expect the government to prove otherwise.

B. This Court should exclude the government's DNA evidence.

Federal Rule of Evidence 702 sets forth the following conditions under which a properly qualified expert can give opinion testimony at trial:

> (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue; (b) the testimony is based upon sufficient facts or data; (c) the testimony is the product of reliable principles and methods; and (d) the expert has reliably applied the principles and methods to the facts of the case.

Fed. R. Evid. 702 (emphasis added).

In *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993), the Supreme Court held that Rule 702 assigns the court a "gatekeeper" role and charges it with the task of ensuring that expert testimony "'rests on a reliable foundation and is relevant to the task at hand.'" *United States v. Hermanek*, 289 F.3d 1076, 1093 (9th Cir. 2002) (quoting *Daubert*, 509 U.S. at 597). This "gatekeeper" role requires the court to assess "whether the reasoning or methodology underlying the testimony is valid and [] whether that reasoning or methodology properly can be applied to the facts in issue." *Id.* (quotation omitted). The Supreme Court clarified in *Kuhmo Tire, Co. v. Carmichel*, 526 U.S. 137, 147 (1999), "that the district court's duty to act as gatekeeper and to assure the reliability of proffered expert testimony before admitting it applies to all (not just scientific) expert testimony." *Hermanek*, 289 F.3d at 1093. That is, Rule 702 "establishes a standard of evidentiary reliability" for *all* such matters. *Kuhmo Tire*, 526 U.S. at 149.

Under *Daubert* and *Kumho Tire*, the court must "make certain that an expert, whether basing testimony upon professional studies or personal experience, employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field." *Kuhmo Tire*, 526 U.S. at 152. To that end, *Daubert* and *Kumho Tire* set forth a non-exhaustive list of factors bearing on the reliability and validity of expert testimony: (1) whether the theory or technique can be and has been tested, (2) whether the theory or technique has been subjected to peer review and publication, (3) the known or potential rate of error, (4) whether there are standards

controlling the technique's operation, and (5) the degree of acceptance within the relevant scientific community. *Daubert*, 509 U.S. at 593–94.

For the reasons stated above, it is impossible without further discovery to address fully whether the government's length-based STR testing meets admissibility standards under *Daubert* and under Federal Rule of Evidence Rules 104(a), 402, 403, 702, and 703, 18 U.S.C. § 3593(c), and the Fifth, Sixth, and Eighth Amendments to the United States Constitution. It is possible, however, to at least offer some preliminary analysis of whether the government's STRmix™ analysis meets these standards.

1) Weak likelihood ratio (LR) statistics using STRmix™ are unreliable and more prejudicial than probative.

STRmix™ is billed as expert forensic software that can "resolve previously unresolvable mixed DNA profiles." *See* STRMIX™, www.strmix.com/strmix/ (last visited Jan. 3, 2021). It claims to be a a fully continuous approach for DNA profile interpretation that is run without the need of high-speed computing on the part of its customers.

According to STRmix™ promotional materials, the program "combines biological modeling and mathematical processes to interpret a wide range of complex DNA profiles." STRMIX™, https://www.strmix.com/strmix/how/ (last visited Jan. 3, 2021). The program purports to use "well-established statistical methods . . . to build millions of conceptual DNA profiles." *Id.* It also purports to grade them against the evidential sample to provide a meaningful comparison. "A range of Likelihood Ratio options are provided for subsequent comparisions to reference profiles. Using the Markov Chain Monte Carlo engine, STRmix™ models any types of allelic and stutter peak heights as well as drop-in

and drop-out behaviour." *Id.* Finally, it claims to do all of this using "traditional methods" and does so in a conservative and scientific manner. *Id.* The problem, however, is that this is not accurate.

Recent disclosures in June 2020 made by ESR—the creators of STRmix™—to their customers (the FBI is one of these customers) revealed that STRmix™ does not function as well as they have previously claimed that it does.[41] The email from ESR disclosed that ESR has discovered that one of the key methods in their program, calculating the highest posterior (HPD) method for applying a lower bound to the variation included by the Monte Carlo effect in the probabilistic genotyping system, does not function as effectively as previously represented. In fact, ESR further disclosed that tests show that rather than the STRmix™ program operating within the desired and scientifically accepted range of 99% confidence, the program rather has a lower bound coverage as low as 76%. In plain language this means that the STRmix™ program's reported LR in a case does not have a low error rate, is not the most conservative LR that may be observed in modeling, and falls well outside of the general scientifically acceptable confidence interval in the field.

Thus, error rates exist within the LR statistics generated in case work well outside of the typical recognized scientific confidence intervals. This renders the actual reported statistical numbers unreliable. Since genotype weights are at the heart of STRmix™ results, before the introduction of any evidence associated with LRs in the present case,

[41] Undersigned counsel learned of the existence of this ESR email outside of the discovery process in this case.

this Court should require the government through the FBI Labortaory, or ESR—the creator of the STRmix™ program—to prove that the LRs are conservative and accurate. This is especially true when the LRs are as weak as the LRs at issue here.

Finally, the characterization of LRs in the single digits and the low double digits on either side of inclusion or exclusion as "weak" rather than "limited" is a representation that is taken directly from Dr. John Buckleton, one of the premier scientists at ESR responsible for the creation of the STRmix™ program. Dr. Buckleton has previously testified that LRs within the range 1000 of either inclusion and exclusion fall within a range that he has previously recognized as "uninformative." He has agreed to the label of these being "weak" and simply numbers that clearly fall within recognized ranges where errors in false inclusions and exclusions are found. Transcript of Hearing, *People v. Anderson*, No. 17CR677 (Colo. Dist. Ct. County of Adams July 19, 2019), at 39–40 (testimony of Dr. Buckleton). If one of the principal creators of the STRmix™ program believes that numbers like those presented in this case are "weak" and are in the range of "uninformative" than this Court should exclude the LRs within those ranges in this case.

2) The "verbal equivalency" language adopted by the FBI DNA Laboratory is unreliable, more prejudicial than probative, and will lead to jury confusion in this case about the strength or weakeness of the DNA evidence.

In 2018, SWGDAM, the FBI-sponsored guideline setting body in the area of forensic DNA typing, published a recommended guide for reporting likelihood ratios with "verbal equivalences." The language for the "verbal equivalencies" was recommended by an Ad Hoc Committee and adopted by the Executive Committee.

SWDGAM COMMUNICATIONS, https://www.swgdam.org/publications (last visited Jan. 4, 2021). The SWGDAM verbal equivalencies are subjective and arbitrary. There is no scientific or empirical support for the verbal equivalencies contained within the suggested terms recommended by SWGDAM. Further, there is a lack of consensus in the field over the verbal equivalence language recommended by SWGDAM. Accordingly, the proposed verbal equivalency statements in Agent Plaza's report must be excluded at trial.

V. CONCLUSION

For the reasons set forth above, and any additional reasons that may arise after full discovery and a hearing on this Motion, Mr. Bowers respectfully moves this Court for an order: (1) compelling the government to provide forthwith the "case litigation" packets that support the findings and conclusions of FBI DNA Casework Analyst Marcy L. Plaza's 16-page DNA Report; (2) granting the defense a reasonable time to review any ordered discovery and to file a supplemental pleading in support of this motion; (3) granting an evidentiary hearing on this *Daubert* motion; and (4) following the holding of such a hearing and the filing of any post-hearing briefing, granting the motion to exclude the government's DNA evidence.

Respectfully submitted,

/s/ Judy Clarke
Judy Clarke
Clarke Johnston Thorp & Rice, PC

/s/ Michael J. Novara
Michael J. Novara
First Assistant Federal Public Defender

/s/ Elisa A. Long
Elisa A. Long
Supervisory Assistant Federal Public Defender

IN THE UNITED STATES DISTRICT COURT
FOR THE WESTERN DISTRICT OF PENNSYLVANIA

UNITED STATES OF AMERICA	)	
	)	
v.	)	Criminal No. 18-292
	)	
ROBERT BOWERS	)	

ORDER

Upon consideration of the Defendant's . . ., it is ordered that the Motion is GRANTED. The government is ORDERED to provide . . . .

_____
Date

_____
Donetta W. Ambrose
United States District Judge

Exhibit 1
(Filed Under Seal)